

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

589 AsnProLysGlnProThrAsnPropheLeuGluMetIleLysPheLeuLe 605
 173 TTCCCTAAGCAGCGCAAAATCCATTCTAGAGATGGTGAATTCCTGCT 222
 605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuI 622
 223 GGAGAGAATTGGCACCAGTCACATGACTGCTGAGGCAATAAGTGCCTGG 272
 622 leLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlu 638
 273 TTAAGCTGTTAATAAGTCAATTAAGGCACTGCTGATGATGAAGATGAG 322
 639 GlyValProThrAspGlnAlaIleAlaGlyLeuGluLeuLeuLysVa 655
 323 GGAGTAACCTGATGACTGCCATCCGCTGCTGGACTTGAGCTACTTAAGGT 372
 655 lleuSerPheThrHisProIleSerPheIleSerAlaGluThrPheGluS 672
 373 TTATCATTTACTCAACCCCTGCGTTCACCTGCGCTGAACATACGAGT 422
 672 erLeuLeuAlaCysLeuLysMetAspGluLysValAlaGluAlaAla 688
 423 CTCTGCTGAGTCTGAAATGGAGGAGCAATAAGTGGCTGAAGCAGCA 472
 689 LeuGlnIlePheLysAsnThrGly 696
 473 ATACAGATCTTCAGGAACACAGCA 496

seq_name: gb_est73:BE306904

seq_documentation_block: 26-OCT-2000
 LOCUS BE306904 557 bp mRNA EST
 DEFINITION 601087447F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3482314 5',
 mRNA sequence.
 ACCESSION BE306904
 VERSION BE306904.1 GI:9161192
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 557)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 P1: L148511 row: n column: 11
 High quality sequence stop: 546.
 Location/Qualifiers
 1. 557
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:3482314"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stages="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

FEATURES
 source

BASE COUNT 160 a 118 c 132 g 147 t

ORIGIN

alignment_scores: Quality: 670.00 Length: 185
 Ratio: 4.012 Gaps: 4
 Percent Similarity: 90.270 Percent Identity: 76.216
 alignment_block:
 US-09-512-581-2 x BE306904 ..
 Align seg 1/1 to: BE306904 from: 1 to: 557
 781 PheAlaAlaProTriPlySerTrpValAlaThrPheIleValLysAspLe 797
 3 TTTGGCTCCCGCATGAATCTGTAGTGGCAAACTTTATTGTTAAAGACCT 52
 797 uLeuMetAsnAspArgLeuProGly..LysLysThrThrLysLeuTrpVa 813
 53 TCTAATGAACGACAGGTCAACAGGTGACGAAGAANTGGAACACTTATGGTC 102
 813 lProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleL 830
 103 TCCAGATGAGGAAGTTCTCTGGAAGTTCTAGCAAAAGGTACAGGCAATTA 152
 830 ysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSer 846
 153 AACTTCTGCTAAGGTGGCTGTTGGGTATGAAAAACAACCAAGTCTAAATCT 202
 847 GlyThrSerThrLeuArgLeuLeuThrIleLeuHisSerAspGlyAs 863
 203 GCCAATCAACCTCTCGGTATTATTCAGCCATGCTGGTTAGTAGGGCGGA 252
 863 pLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgL 880
 253 CCTGCAGACGAAAGAGGATCAGTAAATCTGATATGCTCGCTTGGAT 302
 880 euAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
 303 TAGCTGCTGCTAGTGCATATAAGAGCTTGCTCAGGAACCTTGTACCAT 352
 897 GluIleIleThrLeuGluGlnTyrGlnLeu.CysAlaAlaAlaIleAsnA 913
 353 GAAATATTACCCCAAGACAGTTTCTAGCTACTGTGCACCTGGTTATTATG 402
 913 spGluCysTyrGlnVal.ArgGlnValPheAlaGlnLysLeuHisLysGI 929
 403 ATGAGTGTCTACCAAGTAACGCGCAGATATTTGCCAGAACCTTCATAAGC 452
 929 yLeuSerArgLeuArgLeuProLeuGluThrMetAlaIleCysAlaLeuC 946
 453 ACTTGTGAAGTTGCTTCTCCCACTGGAGTATATGGCCATCTTTGCTTTGT 502
 946 ysAlaLysAspProValLys.GluArgArgAlaHisAlaArgGlnCys 961
 503 GTGCCAAAGACCTCTGTGAACGACAGACAGACACGCTCGGCAGTGT 550

seq_name: gb_est10:AA630616

seq_documentation_block: 391 bp mRNA EST 06-MAR-1998
 LOCUS AA630616
 DEFINITION ac1le07:sl StrataGene HeLa cell s3 937216 Homo sapiens cDNA clone
 IMAGE:856164 3', mRNA sequence.
 ACCESSION AA630616
 VERSION AA630616.1 GI:2553227
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 391)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 WashU Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@watson.wustl.edu
 CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and Ressourcenzentrum Primatenbank, Berlin, Germany (web address: www.rzpd.de)
 Possible reversed clone: polyT not found
 Seq primer: T7 ET from Amersham
 High quality sequence stop: 454.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1. .498
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish WashU MPIMG EST"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield stage embryos"
 /lab_host="X11-blue MRF"
 /note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTTCATGACCGAGCGCCGCTTTTCTTTTCTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 147 a 117 c 121 g 113 t
 ORIGIN

alignment_scores:
 Quality: 673.00 Length: 158
 Ratio: 4.610 Gaps: 0
 Percent Similarity: 92.405 Percent Identity: 81.646

alignment_block:
 US-09-512-581-2 x AI558326 ..

Align seg 1/1 to: AI558326 from: 1 to: 498

539 ProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAs 555
 |||||
 23 CCAGGAAAGCTCAGGACTTTATGAAGAAGTTAATCAGGTTCTGGGTGA 72
 |||||
 555 pAspGluLysIleArgLysGlnLeuValLeuValSerProThrCys 572
 |||||
 73 AGACGAAAGTACGCTCTCAGCTTGAGCAACTCATCAGCCCACTGCT 122
 |||||
 572 erCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGly 588
 |||||
 123 CATGTAAACAGCGCCGACAGTGTGTGAGGGAATACACACGAAAGCTGACA 172
 |||||

seq_name: gb_est21:AI558326

seq_documentation_block:
 Locus AI558326 498 bp mRNA EST 24-MAR-1999
 DEFINITION fb78906.x1 zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA sequence.
 ACCESSION AI558326
 VERSION AI558326.1 GI:4508564
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
 1 (bases 1 to 498)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,

alignment_scores:
 Quality: 677.00 Length: 165
 Ratio: 4.340 Gaps: 0
 Percent Similarity: 94.545 Percent Identity: 73.939

alignment_block:
 US-09-512-581-2 x BE925887 ..

Align seg 1/1 to: BE925887 from: 1 to: 529

415 IleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLy 431
 ::::|
 35 CTTTATAGAAATACTGCTCGGTGAGCAGGAAGGAAGCTGCGAGA 84
 ::::|
 431 sGlnIleAlaTPrIleLysAspLysLeuLeuHisIleTyrTyrGlnAsn 448
 ::::|
 85 GAAAGTCAGCTGGATAAAGGACAACTTCGCATATTATTATCAGAAC 134
 ::::|
 448 erIleAspAspArgLeuValGluArgIlePheAlaGlnTyrMetVal 464
 ::::|
 135 GCATTGACGACAACTGCTGCTAGAGAAATCTTTGCTCAGTACTTGT 184
 ::::|
 465 ProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 481
 ::::|
 185 CCCCAACCTGGAAGAGAGAGAGAGAAATGCTTATATTACTTATA 234
 ::::|
 481 rAlaThrLeuAspLeuAsnAlaValLysAlaLeuAspGluMetPrLysC 498
 ::::|
 235 TGCTAGTTGGATCCAACTGCTGTAAGAGCTCTCAAGAAATGCGAGT 284
 ::::|
 498 ysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLys 514
 ::::|
 285 GTCAGAACATGCTTCGGAGGCTATGCGGAACCTATTGGATTGCAACA 334
 ::::|
 515 GlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMetVa 531
 ::::|
 335 CAGCTCATACAGAGGCTACTGCTGCGCATGTTGGAAACATGATGAC 384
 ::::|
 531 lIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysL 548
 ::::|
 385 CATAGCAAGAAATTTGCTGACCCCGGGAAGACACAGATTTTGTGA 434
 ::::|
 548 ysPheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGlu 564
 ::::|
 435 AATTAAACAGGTTCTCGGCGATGATGAGAAACTTCGCTCAGTTGGAG 484
 ::::|
 565 ValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCys 579
 ::::|
 485 TTATTATTAGCCCAAGCTGTTCTTGAACACAGCAGATATTGTTGT 529
 ::::|

COMMENT

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://ww.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html
Plate: 158 row: H column: 23
Seq primer: SP6
Class: BAC ends.

FEATURES

source

BASE COUNT	290 a	148 c	163 g	150 t	
ORIGIN					
alignment_scores:					
Quality:	677,50	Length:		206	
Ratio:	4,234	Gaps:		2	
Percent Similarity:	77,670	Percent Identity:		67,961	

354 AGACATTAAGTCAAAAGACAAACAGACGACCCACATCAAGAGAGACACAAC 403

1354 lnArg.....ProGluSerSerAlaIleGluSerThrGlnS 1355

404 AGAGGTGAGTGTGTGCAGCTCTAAAGTGGACTTCGGTGCAGTAAATGTGC 453

1355AGTGCAGTTTGAATCACACAGT 552

1369 erThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSer 1385

553 CCACACCACAGAAAGGACGAGGAGACCATCAAAAGCACCCTCACCATCA 602

1386 GlnProLysLysAsnVal 1391

603 CAACCCCCCAAAAAAATG 620

seq_name: qb_est81:BE925887

seq documentation block:

LOCUS	BE925887	529 bp	mRNA	EST	02-OCT-2000
DEFINITION	QV3-BN0047-300800-278-h04	BN0047	Homo sapiens	cDNA, mRNA sequence.	
ACCESSION	BE925887				
VERSION	BE925887.1	GI:10451963			
KEYWORDS	EST,				
SOURCE	human.				

SOURCE ORGANISM

ORGANISM HOMO SAPIENS
Eukaryota: M

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates: Catarrhini: Hominoidea: Homo.

REFERENCE 1 (bases 1 to 529)

AUTHORS
Dias Neto, E., Garcia

bras negro, L., *Carta Corica*, N., *Veljovski gimnazi*, 35, *Belones*, 1911, N.
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and

Sample ID	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Sample ID	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT
Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwlg.org.br/scrpt>)

800-278-h04&t3=2000-08-30&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 34

High quality seq

FEATURES

Location/Qualifiers

source 1. .529

```
/organism="Homo sapiens"
```

/db_xref="taxon:96

```
/clone_lib="BN0047"
```

```
/dev_stage="Adult"
```

```

/note="Organ: breast_normal; vector: puc18; site_1: SmaI;
site_2: SmaI. A small library was made by cloning products

```

Site_2: SmaI; A mini-library was made by cloning products derived from *ORF2* PCR in a *Pst*I-SmaI vector.

derived from ORESTES PCR (U.S. Letters patent application No. 106 716 - Ludwig Institute for Cancer Research)

No. 196, /16 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector Reverse transcription of

profiles into the pUC 18 vector tissue mRNA and cDNA amplifi-


```

seq_documentation_block:
LOCUS      BF107038             818 bp      mRNA           EST           19-OCT-2000
DEFINITION 601824203F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043905 5',
            mRNA sequence.
ACCESSION  BF107038
VERSION     BF107038.1 GI:10889663
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 818)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            CDNA Library Preparation: CLONTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM867 row: f column: 02
            High quality sequence stop: 630.
            Location/Qualifiers
              1..818
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4043905"
                /clone_lib="NIH_MGC_79"
                /lab_host="pH108 (T1 phage-resistant)"
                /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                Site_1: SfII (ggccattatggcc); Site_2: SfII (ggccattatggcc)
                ); 5' and 3' adaptors were used in cloning as follows: 5'
                adaptor sequence: 5'-CAGGGCCATTATGGCC-3' and 3' adaptor
                sequence: 5'-ATTCTAGAGCGGCGGCGGCGAGATG-dt(30)BN-3'
                (where B = A, C, or G and N = A, C, G, or T). Average
                insert size 1.3 kb (range 0.5-4.0 Kb). 15/15 colonies
                contained inserts by PCR. This library was enriched for
                full-length clones and was constructed by Clontech
                Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                Library."
BASE COUNT 251 a 173 c 185 g 209 t
ORIGIN
alignment_scores:
  Quality: 678.50      Length: 274
  Ratio: 3.043
  Percent Similarity: 81.387      Percent Identity: 61.679
alignment_block:
us-09-512-581-2 x BF107038
Align seg 1/1 to: BF107038 from: 1 to: 818
460 AlaGlnTyrMetValProHisAsnLeuGluThrGluArgMetLysC 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 GCTCAGTATCTGTGCCCAACCTGGAACAGAGAGAGAGAGAAATG 68
476 sLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuA 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 CTTATATTACTTATATCTAGTTGGATCAAAATGCTGTAAGAGCTCTCA 118
493 snGluMetTTrpLysCysGlnAsnLeuArgHisGlnValLysAspLeu 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 ACGAAATGTGGAAGTGTCAAGACATGCTTCGGACCAATGTACGGCAA... 165
510 Leu.AspleuLeuLysGlnProLysThrAspAlaSerValLysAlaLeuP 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

166 CTATGATTGCACAAGCAGCGCTACATCAGAGCGCTAACTGTCTGCCATGT 214
526 heserLysValMetValIleThrArgAsnLeuProAspProGlyLysAla 542
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 T.GGAAACATGATGACCATGCAAGAAATTTGCCTGACCCCGGGAAGCA 264
543 GlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysII 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 CAAGA.TTTCTGGAAGAATTTAACCAGGTCTCGCGCATGATGAGAAACT 313
559 eaArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnA 576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 TCGGTCTCAGTTGGAGTATTAAATTAGCCCAACCTGTTCTTGCACAAAG 363
576 laGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGln 592
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 CAGATATTGTGTGAGAGAAATAGCCCGAAACTTGCAAATCCTTAAGCAA 413
593 ProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAl 609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 CCAACAATCCTTTCTAGAGATGCTCAATTTCTGTG.GAAACAATCGC 462
609 aProValHisIleAspThrGluSerIleSerAlaLeuIleLys.GlnVal 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 ACCTGTGCACATTGATTGAGAAGCCATAAGTGCACACTAGTGAATTTGATG 512
626 AsnLysSerIleAspGlyThr.AlaAspAspGluAspGluGlyValProT 642
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 AATAAGTCAATAGAGGGGACAGGAGATGATGAAGAGAGGGGTAAACGT 562
642 hrAspGlnAlaIleArgAlaGlyLeuLeuLeuLysValLeuSerPhe 658
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 CCAGATACAGATATCCGTTAGGACTTGAATTTCTTAAGTTCTGTC.TTT 611
659 ThrHisProLysSerPheHisSerAlaGluThrPheGluSerLeuLeuAl 675
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
612 ACACATCCTACCTCGTTC...ACTCTGCAGAGACTATGAGTCTCTTACA 658
675 acLysLeuLysMetAspAspGluLysValAlaGlu.....AlaAlaLeuG 690
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 GTGCCTAAAG.....GGGTGCAGGTGGCGAAGTGTATCCACTTTTA 702
690 InIlePheLysAsn...ThrGlySerLysIleGluGluAspPheProHisI 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
703 CGATTCGCTCCAAATGACGACTTTCGATAGATGACTTATTCCTTTC 752
706 leArgSerAlaLeuLeuProValLeuHisLysSerLysLysGlyPro 722
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
753 TAAAGCA.....AGGGGTCCC 769
723 ProArgGlnAlaLys 727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
770 CCACCACAGGGGCGT 784
seq_name: gb_gss34:AZ902024
seq_documentation_block:
LOCUS      AZ902024             751 bp      DNA           GSS           05-MAR-2001
DEFINITION RPCI-24-158H23.TJ RPCI-24 Mus musculus genomic clone RPCI-24-158H23
            , DNA sequence.
ACCESSION  AZ902024
VERSION     AZ902024.1 GI:13220969
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
            1 (bases 1 to 751)
            Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levis,M.,
            Tsengaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebreyegorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Unpublished (1999)

```


alignment_block:

US-09-512-581-2 x AA939593

Align seg 1/1 to: AA939593 from: 1 to: 486

1143 LysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSe 1159
 2 AAACAGTCTCAGACCAAAATCATCAAGAATGGAAACTGTGAGCAACGCAAG 51
 1159 rSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspS 1176
 52 CAGCAGCTCCACCCCAAGCTCTCTGGAAGGATCAAGGGAGGCTTGATA 101
 1176 exSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerPro 1192
 102 GCTCTGAATGGATCAGATGAGAAATGAAGATTATACAAATGCTTCACCT 151
 1193 LeuProGlyLysLysSerAspLysArgAspSerAspLeuValArgSe 1209
 152 TTGCCAGGAAAAAAGTCACAGAGAGAGAGACCCCTGATCTT.....TC 195
 1209 rGluLeuGluLysProArgGlyArgLysLysThrProValThrGluGlnG 1226
 196 TGAGTTGGAGAGCCCTAGAGTCGGAAAAAGACCTGTCACAGCCCTG 245
 1226 luGluLysLeuGlyMetAspLysLeuThrLysLeuValGlnGlnLys 1242
 246 AAGAGAAATAGGTATGATGATGACCTAACTAAGTTGGTACAGGAACAGAAA 295
 1243 ProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSe 1259
 296 CCTAAAGCAGTCAGGAGCGGCGGAAAGAGCGGTACAGCTCAGACTC 345
 1259 rAspGluGlnGlnTrpProGluLysArgLeuLysGluAspIleLeuG 1276
 346 AGACGAGCAGCAGTGCCTGAGGAGAGAGGACCAAGAGGAGCTCTCG 395
 1276 luAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArg 1292
 396 AAAATGAGGATGAGCAGACAGACCCACCAAAAAAGGCAAGAGAGCAGG 445
 1293 ProProLysProLeuGlyGlyThrProLysGlu 1304
 446 CA.CCAAAACCTCTTGGTGGGGGAGCAGCAAGGAAG 480

seq_name: gb_est88:BF451878

seq_documentation_block:

LOCUS BF451878 580 bp mRNA EST 29-DEC-2000
 DEFINITION uz80904.y1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3675414 5' similar to TR:Q9Y451 Q9Y451 ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN. ;, mRNA sequence.
 ACCESSION BF451878
 VERSION BF451878.1 GI:11518047
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 580)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/lresources.shtml

MGI:1436182

Seq primer: -40RP from Gibco

High quality sequence stop: 376.

FEATURES

source

1..580
 /organism="Mus musculus"
 /strain="C57BL/6J (f1)"
 /db_xref="taxon:10090"
 /clone="IMAGE:3675414"
 /clone_lib="NCI_CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 Stem cell origin.
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 161 a 111 c 161 g 147 t

ORIGIN

alignment_scores:

Quality: 680.50 Length: 159
 Ratio: 4.629 Gaps: 1
 Percent Similarity: 92.453 Percent Identity: 83.648

alignment_block:

US-09-512-581-2 x BF451878

Align seg 1/1 to: BF451878 from: 1 to: 580

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrThrProPr 17
 104 ATGGCTCATTCACAGACAGACCAACGATGGGAAATATTACCTCC 153
 17 OGlyValLysGluIleSerAspLysLysSerLysGluGluMetValArgA 34
 154 TGGAGTCAAGGAAATCTCAGATAAAATCTCTAAAGAGAGAGATGGTGAGC 203
 34 rGluLysMetValLysThrPheMetAspMetAspGlnAspSerGlu 50
 204 GGTAAAGATGGTTGTAAAACTTTCATGGACATGGACCCAGGACTCTGAA 253
 51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67
 254 GAGGAAAGGAACTTATCTAAACCTAGCTTTACATCTTCTCTCCTCCT 303
 67 ePheLeuLysHisProGlyLysAspValArgLeuValAlaCysCysL 84
 304 CTTCCTCAAGCATCTCTGATAAAGATGTTCTGTTACTGGTGGCTGCTGCC 353
 84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
 354 TTGCTGATATTTTCAGGATTTATGCTCTGAGGCTCCTTACAGCTCTGCA 403
 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117
 404 GATAGACTTAAGGATATATGTTTATATAACAAGGCGACTGAAGGACT 453
 117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134
 454 AGAAGATACACAGAGCCCTCGATTTAAATTTAGGATTTGTATGATGAG 503
 134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150
 504 GCATGGCTGGGTAAATCATATACATGCTGTTGGAGTTGGAGAGCGC 553
 151 AsnGlu.....IlePheThrGln 156
 554 ATGGAATGCTTACTCATCTTTACAGAG 580

seq_name: gb_est83:BF107038

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 768 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 411.

FEATURES

Location/Qualifiers

1. .483

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1525851"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19w, testis NHT, and B-cell
NCL-GAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT

149 a 101 c 78 g 155 t

alignment_scores:

Quality: 717.00 Length: 158
Ratio: 4.717 Gaps: 0
Percent Similarity: 96.203 Percent Identity: 84.810

alignment_block:

US-09-512-581-2 x AA913927/rev ..

Align seg 1/1 to reverse of: AA913927 from: 1 to: 483

250 AspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrrAsnIleAs 266
|||||
483 GATTGTCAGAACATGATTGATGCTGATTGAGGAACCTTTTGGCTATAGA 434
266 pSerHisLeuLeuSerValLeuProGlnLeuGluPheLeuLys 283
|||||
433 TCCTCATTTATTATTCGCTCATGCGACACAGCTTGAATTCAAACTAAGA 384
283 erAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMet 299
|||||
383 GCAATGATGAGGAAGAGGATTAGCTTCTTGGCTTCAGTTCAGTAAATG 334
300 PheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpG1 316
|||||
333 TTGCGTCCAAAGATTCTGATTTGGCAACACAGAACTGCTCTTTGGCA 284
316 nCysTyLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluC 333
|||||
283 ATGTTTCTTGGACGATTAAATGATATTCATGTTCTCGTGCAGATTAGAAA 234
333 yValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys 349
|||||
233 GTGTGAATTTGCCAGTCATTGTTTAAATGAATCACCAGATTAGCGAAG 184
350 AspLeuThrGluTyLeuLysValArgSerHisAspProGluGluAlaI1 366
|||||
183 GATCTCACAAATATTAAAGTTAGATCATCATGATCCAGAAAGCTAT 134
366 eArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleL 383
|||||

133 TCGTCATGATGTCATCTTACTATAATAACAGCTGCCAAGAGGACCTGG 84

383 euleuValAsnAspHisLeuAsnPheValArgGluArgThrIleuAsp 399

83 CCTTAGTAATATGATGACCTGCTTGGCTTTGTAGGGAAGAACACTGGAT 34

400 LysArgTrpArgValArgLysGlu 407

33 AACCGTGGCGAGTAAGAAAAA 10

seq_name: gb_est13:AA939593

seq_documentation_block:

LOCUS AA939593 486 bp mRNA EST 01-MAY-1998
DEFINITION Y51f06.r1 Soares_thymus_2NbMT Mus musculus cDNA clone
IMAGE:1330019 5', mRNA sequence.

ACCESSION AA939593

VERSION AA939593.1 GI:3100370

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 486)

REFERENCE

AUTHORS

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Mortis, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 485.

MG1:689563

Location/Qualifiers

1. .486

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:1330019"

/clone_lib="Soares_thymus_2NbMT"

/sex="male"

/tissue_type="Thymus"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAGTGGAGCGCCCGCTTTTCTTTTCTTTTCTTTTCTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. RNA

provided by Dr. Bertrand Jordan. Library went through two

rounds of normalization, and was constructed by Bento

Soares and M.Fatima Bonaldo."

BASE COUNT 179 a 103 c 133 g 71 t

ORIGIN

alignment_scores:

Quality: 717.00 Length: 162

Ratio: 4.626 Gaps: 1

Percent Similarity: 95.679 Percent Identity: 88.889

393 AAGAGCCCTCAATTAAATAGGTATTTTATTACTTTGAGAACATTCGAT 344
 137 rpVallysSerTyAsnIleCysPheGluLeuGluAspSerAsnGluIle 153
 343 GGGTTAAATCATATACATATGCTTTGAGTTAGAGATAGCAATGAAATC 294
 154 PheThrGlnLeuTyArgThrLeuPheSerValIleAsnAsnGlyHisAs 170
 293 TTTACTCAACTTTTACAGAACATATTCTCAGTTATTAACAATGGCCACAA 244
 170 nGlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCysG 187
 243 TCAGAAAGTTTATATGACATATGCTGAGCTCATGAGCTCTATCATTTG 194
 187 luGlyAspThrValSerGlnGluLeuLeuLeuAspThrValLeuValAsnLeu 203
 193 AAGGTATACCGTATCTCAGGAGCTCTTAGATACAGTTTTTAGTAAATCTG 144
 204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyAspLeuAlaLysAl 220
 143 GTACCTGCCCATGAAGAACTTAAACAAGCAAGCATATGATTGGCAAGGC 94
 220 aleuLeuLysAsnGlnAlaGlnAlaIleGluProTyIleThrPheP 237
 93 TTTGCTCAAGAGACTGCTCAAGCTATTGACCATATATATACCAATTTT 44
 237 heAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249
 43 TTAACCAAGGTTCTGATGCTTGGGAAAAAAATTCGAGC 6

seq_name: gb_est72:BE252186

seq_documentation_block: 546 bp mRNA EST 13-JUL-2000
 LOCUS BE252186 601113791F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3354263 5',
 DEFINITION mRNA sequence.
 ACCESSION BE252186
 VERSION BE252186.1 GI:9123222
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC http://mhc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LUCM156 row: f column: 24
 High quality sequence stop: 535.
 Location/Qualifiers
 1..546
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3354263"
 /clone_lib="NIH_MGC_16"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

FEATURES
 source

BASE COUNT 181 a 100 c 127 g 138 t
 ORIGIN
 alignment_scores:
 Quality: 726.00 Length: 181
 Ratio: 4.197 Gaps: 0
 Percent Similarity: 95.580 Percent Identity: 76.243
 alignment_block:
 US-09-512-581-2 x BE252186 ..
 Align seg 1/1 to: BE252186 from: 1 to: 546
 358 ArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIl 374
 3 AGATCATCATGATCCAGAGAAGCTATTTCGTCATGATGTCATTGTTACTAT 52
 374 eValThrAlaIleLysLysAspIleLeuValAsnAspHisLeuLeuA 391
 53 AATAACAGCTGCCAGAGAGGACCTGGCCTTAGTAAATGATCAGCTGCTTG 102
 391 snPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGlu 407
 103 GCTTTGTAAGGGAAGAACACCTGGATAAACGGTGGCGAGTAAGAAAGAA 152
 408 AlaMetMetGlyLeuAlaGlnIleTyLysLysTyAlaLeuGlnSerAl 424
 153 GCTATGATGGTCTGCTCAGCTTTATAGAAATACTGCTTCATGGTGA 202
 424 alaGlyLysAspAlaIleLysGlnIleAlaTrpIleLysAspLysLeuL 441
 203 AGCAGGAAAGAGAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAACTTC 252
 441 euHisIleTyArgGlnAsnSerIleAspAspArgLeuLeuValGluArg 457
 253 TGCATATTTATTATCAGAACAGCATTTGACGACAACTGTTGGTAGAGAA 302
 458 IlePheAlaGlnTyMetValProHisAsnLeuGluThrThrGluArgMe 474
 303 ATCTTTGCTCAGTATCTTCTCCCAACACCTGGAAACAGAGAGAAT 352
 474 TyLysCysLeuTyArgThrLeuAlaThrLeuAspLeuAsnAlaValLysA 491
 353 GAATGCTTATATATCTATATGCTAGTTGGATCCAAATGCTGTAAAG 402
 491 laLeuAsnGluMetTyrLysCysGlnAsnLeuLeuArgHisGlnValLys 507
 403 CTCTCAACCAATGTGGAAGTGTCAAGACATGCTTCGGAGCCCATGTACGC 452
 508 AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAl 524
 453 GAATATTGATTTGTCACAAAGCAGCCTACATCAGAGGCTAACTGTTCGTC 502
 524 allePheSerLysValMetValIleThrArgAsnLeuProAsp 538
 503 CATGTTT.GGAAAACTGATGACCATAGCAAGAAATTTTCCTGAA 544

seq_name: gb_est13:AA913927

seq_documentation_block:
 LOCUS AA913927 483 bp mRNA EST 10-JUN-1998
 DEFINITION O139f02.s1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
 IMAGE:1525851 3', mRNA sequence.
 ACCESSION AA913927
 VERSION AA913927.1 GI:3053319
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 483)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: Not; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

BASE COUNT 50 a 100 c 90 g 183 t
ORIGIN

alignment_scores:
Quality: 731.00 Length: 140
Ratio: 5.221 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-512-581-2 x AI680124/rev ..

Align seg 1/1 to reverse of: AI680124 from: 1 to: 423

```
1247 GlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnG1 1263
|||||
422 CAGGAAGTCGGAAGAGAGGCCATACGGCTTCAGATCTGATGACAGCA 373
|||||
1263 nTTPProGluGluLysArgLeuLysGluAspPileLeuGluAsnGluAspG 1280
|||||
372 GTGGCCTGAGGAAGAGCGCTCAAGAAGATATATTAGAAAATGAAGATG 323
|||||
1280 LuGlnAsnSerProProLysLysGlyLysArgGlyArgProProLysPro 1296
|||||
322 AACAGATACTCGCCCAAAAGGGTAAAGAGCGCCAGCACCAACCAACCT 273
|||||
1297 LeuGlyGlyThrProLysGluGluProThrMetLysThrSerLysLy 1313
|||||
272 CTGTGCGAGGTACACCAAGAGAGAGCCCAACATGAAACTTCTAAAAA 223
|||||
1313 sGlySerLysLysLysSerGlyProProAlaProGluGluGluGluG 1330
|||||
222 AGGAAGCAAAAAAATCTGGCTCCAGCACCAGAGGAGGAGGAAG 173
|||||
1330 LuGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHis 1346
|||||
172 AAGAAAGACAAAGTGGAAATACGGAACAGACAGAGTCCAAAGCAACAGC 123
|||||
1347 ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAl 1363
|||||
122 CGAGTGTCAAGGAGAGCAGACAGAGAGAGATCTCTCTGAATCTAGTGC 73
|||||
1363 alleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerL 1380
|||||
72 AATTGAATCCACACAGTCCACACACAGAAAGGAGGAGGAGGAGGAGGAG 23
|||||
1380 ysthrProSerProSerGln 1386
|||||
22 AAAGCCATCACCATCACAA 3
```

seq_name: gb_est81:BE948523

seq_documentation_block:
LOCUS BE948523 445 bp mRNA EST 03-OCT-2000
DEFINITION UI-M-BH3-awr-g-03-0-UI-s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-awr-g-03-0-UI 3', mRNA sequence.
ACCESSION BE948523
VERSION BE948523.1 GI:10526282
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 445)
AUTHORS Bernaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE
COMMENT

97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.

FEATURES

source

Location/Qualifiers
1..445

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-awr-g-03-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bernaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_SEQ=None found"

BASE COUNT 134 a 83 c 74 g 154 t
ORIGIN

alignment_scores:

Quality: 731.00 Length: 146
Ratio: 5.076 Gaps: 0
Percent Similarity: 98.630 Percent Identity: 97.260

alignment_block:

US-09-512-581-2 x BE948523/rev ..

Align seg 1/1 to reverse of: BE948523 from: 1 to: 445

```
104 LysAspPilePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspTh 120
|||||
443 AAGATATATTTATGTTTAAACAGGCACTAAGGAGCTAGGAAGATAC 394
|||||
120 rLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlat 137
|||||
```



```

DEFINITION BP250017A10D2 Soares normalized bovine placenta Bos taurus cDNA
clone BP250017A10D2 5', mRNA sequence.
ACCESSION BF043498
VERSION BF043498.1 GI:10760553
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 595)
AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinaz,J., Liu,L. and Larson
,J.H.
TITLE Bovine ESTs
JOURNAL Unpublished (2000)
COMMENT Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Trimmi g:
Cross match from Washington University Genome Center PHRAP suite.
This sequence is vector free and at least 200 bp in length.
PCR primers
FORWARD: TATACGACTCACTATAGG
BACKWARD: ATTAACCCCTCAATAAG
Insert Length: 595 Std Error: 0.00
Plate: BP250017A10 row: D column: 2
Seq primer: AGCGATAACAATTTACACAGGA
High quality sequence stop: 595.
FEATURES             Location/Qualifiers
     source
       1..595
         /organism="Bos taurus"
         /db_xref="taxon:9913"
         /clone="BP250017A10D2"
         /sex="female"
         /lab_host="DH10B"
         /note="Organ: placenta; Vector: pT73pac; Site:1: EcoRI;
         Site:2: NotI; The cDNA library was contributed by the
         Soares laboratory and it was constructed and normalized
         as described by Bonaldo, M.F., Lennon, G. and Soares,
         M.B. (1996), Genome Research 6(9): 791-806."
BASE COUNT      249 a 110 c 151 g      85 t
ORIGIN
alignment_scores:
  Quality: 739.50      Length: 169
  Ratio: 4.593      Gaps: 1
Percent Similarity: 95.266      Percent Identity: 85.207
alignment_block:
US-09-512-581-2 x BF043498      ..
Align seg 1/1 to: BF043498 from: 1 to: 595
1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrPr 1221
2 GATCTTGTAAGTCTGAATTGGCAACCAAGAGCCGCAAGAACACACC 51
1221 oValThrGluGlnGluGluLysLeuGlyMetAspLeuThrLysLeuV 1238
52 TGTAACAGATTCAGAGAGAAATAGGCATGTGATGACCTTTCAGTTGG 101
1238 aLcGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHis 1254
|||||

```

```

102 TACAGGAACAGAAACCTAAAGCAGCTACCGTGGAGCAAGAGAGGCCCAT 151
1255 ThrAlaSerGluSerAspGluGlnGlnTPrProGluGluLysArgLeuLy 1271
152 ACAGCTTCAGAGTCAGACCAACAGCAGTGGCTGAGGAAAGAGGCTCAA 201
1271 sGluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysG 1288
202 AGAAGATATATAGAAAATGAGGATGACAGACACAGCTCCACCAAAAAAG 251
1288 lYlYsArgGlyArgProProLysProLeuGlyGlyGlyThrProLysGlu 1304
252 GTAAAGAGGCGCCGACCAACCTCTTGGTGGGGGTACACCAAGAA 301
1305 GluProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyPr 1321
302 GAGCCAACGATCAAAACATCCAAAAAGGAGCAAAAAAATCTCGACC 351
1321 oProAlaProGluGluGluGluGluGluGlnSerGlnSerGlyAsnThrG 1338
352 TTCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
1338 luGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGln 1354
402 AGCAGAAATCAAAAAGTAACAGCAGCCGAGCTTCAAGGAGAGAGAGAG 451
1355 ...ArgAlaGluSerProGluSerSerAlaIleGluSerThrGlnSerTh 1370
452 AGCCGTGTAGGAGCGCTCTAAACAGCAGCTACTAAAGAAAATGACTCAAG 501
1370 rProGln 1372
502 TGAAGAA 508
seq_name: gb_est23:AI680124
seq_documentation_block:
LOCUS AI680124 423 bp mRNA EST 15-DEC-1999
DEFINITION tw65b01.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2264521 3',
mRNA sequence.
ACCESSION AI680124
VERSION AI680124.1 GI:4890306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 423)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1158 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source
       1..423
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:2264521"
         /clone_lib="NCI_CGAP_Ut3"
         /tissue_type="poorly-differentiated endometrial
         adenocarcinoma, 2 pooled tumors"

```



```
1010 AspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuG1 1026
|||||
3 GATATTGAACAACCTAGACAGCTTAAGAGAGTCTTTGGTTGCTTGA 52
|||||
1026 uileLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysM 1043
|||||
53 AATACTAATGCTAAAAATGAAATAATAGTCATGCGCTTTATCAGAAGA 102
|||||
1043 etValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAla 1059
|||||
103 TGTFAGAAAAATATTAACAACAAGAGATGCTCAAGGACGAGATGATGCA 152
|||||
1060 LysMetAsnGluLysLeuTyrThrValLysAspValAlaMetAsnIleI1 1076
|||||
153 AAAATGAATGAAAACTGTATCTATGTATGTATGCTTCCCATGATATCAT 202
|||||
1076 eMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValL 1093
|||||
203 CATGTCAAAAAGACCAACATATAGTTTGGAAATCTCCTAAAGACCGGTAC 252
|||||
1093 euProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLys 1109
|||||
253 TGCAGCTCGCTTTTTCACCCAGCGCTCACAAGAACTTCAGTAACACCAA 302
|||||
1110 AsnTyrLeuProGluMetLysSerPhePheThrProGlyLysProLy 1126
|||||
303 AATATCTGCTCCAGAAATGAATCATTTTCTACTCTCTGGAAACCCAA 352
|||||
1126 sThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyL 1143
|||||
353 AACAAACCAATGTTCTAGGAGCTGTTAATAAGCGCTTTCATCAGCAGGCA 402
|||||
1143 ysGlnSerGlnThrLysSerSerArgMetGlu 1153
|||||
403 AACAACTCAGACAAATATCATCAGAAATGGA 434
```

seq_name: gb_est8:AA511220

```
seq_documentation_block: 437 bp mRNA EST 08-JUL-1997
LOCUS v556b10.r1 Soares_mammary_gland_NBMGM Mus musculus cDNA clone
DEFINITION IMAGE:890971 5', mRNA sequence.
ACCESSION AA511220
VERSION AA511220.1 GI:249074
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 437)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geiseli,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Willson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:518931
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 407.
Location/Qualifiers
i. .437
/organism="Mus musculus"
```

```
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:890971"
/clone_lib="Soares_mammary_gland_NBMGM"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 117 a 96 c 111 g 113 t
ORIGIN
```

```
alignment_scores:
Quality: 741.00 Length: 145
Ratio: 5.110 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.310
alignment_block:
US-09-512-581-2 x AA511220 ..
Align seg 1/1 to: AA511220 from: 1 to: 437
```

```
857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysLeuSerLysPr 873
|||||
3 ATACTGCATAGTGTATGGGATTTGACAGAAACAAGGAAATATTAGTAACC 52
|||||
873 oAspMetSerArgLeuAlaAlaGlySerAlaIleValLysLeuA 890
|||||
53 AGATATGTCAGCGCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGG 102
|||||
890 laGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeu 906
|||||
103 CACAGGAGCCCTGTTTACCACGAGATCATTTACACTGGAGCAGTACCAGCTG 152
|||||
907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAl 923
|||||
153 TGTGCATTAGCCCATCAATGATGAGTGTATCAAGTCAGGCAGGTGTTCCG 202
|||||
923 aGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrM 940
|||||
203 TCAGAAACTTCACAAAGAGCCCTTCCCGCTTACGGCTTCCCTTCCCTGAGTACA 252
|||||
940 etAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
|||||
253 TGCCCATCTGTGCTCTTTGTGCCAAAGACCCCTGTGAAAGAGAGCGAGCC 302
|||||
957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLe 973
|||||
303 CATGCTAGACAGTGTGCTGGTGAAGAACATCATCTGTGAGGAGGAGTACCT 352
|||||
973 uLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluT 990
|||||
353 GAAGCAGCATGCAGCTGTTAGTGAATAATATTGTCCTCTTCTACCAGAGT 402
|||||
```

seq_name: gb_est82:BF043498

```
seq_documentation_block:
LOCUS BF043498 595 bp mRNA EST 10-OCT-2000
```


ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE917453
BE917453.1 GI:10419093
EST
house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 944)

NIH-MGC <http://mgs.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM9134 row: i column: 14

High quality sequence stop: 661.

Location/Qualifiers

1..944

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3964501"

/clone_lib="NCI_CGAP_Maml"

/tissue_type="tumor, biopsy sample"

/dev_stage="3 months, virgin"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 290 a 194 c 246 g 214 t

ORIGIN

alignment_scores:

Quality: 751.00 Length: 252

Ratio: 3.308 Gaps: 2

Percent Similarity: 90.079 Percent Identity: 66.270

alignment_block:

us-09-512-581-2 x BE917453 ..

Align seq 1/1 to: BE917453 from: 1 to: 944

```

493 AsnGluMetRplysCysGlnAsnLeuLeuArgHisGlnVallysAspLe 509
|||||
2 AATGAATGTGGAATGTCAGACATGCTCGAAGTCATGTGCGTGAAC 51
|||||
509 uLeuAspLeuLeuGlnProLysThrAspAlaSerValLysAlaIleP 526
|||||
52 ATG.GACTTACACAGCAGCCTACATCAGAGCGAAGTGTCTGCCATGT 99
|||||
526 heSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAla 542
|||||
100 T.GGGAACTGATGACCATACCAAGAAATTTGCTGACCTGGAAAAGCA 149
|||||
543 GlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysI 559
|||||
150 CAAGA.TTTTAAAGAAATTTAACCAGGTCTTGGTGATGATGAGAACT 198
|||||
559 eArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGln 576
|||||
199 GAGGTCTCACTTGAATATTATATCAGCCCAACCTGTTTCATGCAAGCAG 248
|||||
576 laGluGlyCysValArgGluIleThrLysLysLysLeuGlyAsnProLysGln 592
|||||
249 CTGACGT.TGTGTGAGGAAATATGTCGAAACTTGCAGAACTTGCAGTAA 297

```

```

593 ProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAl 609
|||||
298 CCAACCAACCCCTTTCTAGACATGGTCAAATTTCT.TTGGAAAGAAATGC 346
|||||
609 aProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVal 626
|||||
347 TCCTGTGCACATCGATTTCAGAACCCATAAGTCAGCTGGTAAACTGATGA 396
|||||
626 snLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThr 642
|||||
397 ATAAGTCAATCGAAGGACGCGCGATGATGAGAGAGAGGCGTGCAGTCCA 446
|||||
643 AspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheTh 659
|||||
447 GACTACGCCATCGCTCAGGACTTGAGCTTCTTAAGGTTCTGTCTTTCAC 496
|||||
659 rHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaC 676
|||||
497 ACATCTACCTCGTTCACCTGCAGAGACATATGAGTCTTGTTCACAGT 546
|||||
676 ystLeuLysMetAspAspGluLysValAlaGluAlaLeuGlnIlePhe 692
|||||
547 GCCTAAGATGGAGGATGACAAGGTAGCAGTAGCAAGCTGCGACATACAA 596
|||||
693 LysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAl 709
|||||
597 AGAACACAGCAGCAAAATAGAAACTGAGCTTCCCCAGATACGCTCCAC 646
|||||
709 aLeuLeuProValLeuHisLysSerLysLysGlyProArgGlnA 726
|||||
647 TTGATCCCCA...TTTATCATCAGAACCAAGGGGGGAGCTCACAAACCAAG 693
|||||
726 laLysThrAlaIleHisCysIleHisAlaIlePheSerLysGluThr 742
|||||
694 CAAGAAGTGT...CACTGGATTTCATGTCATCTTCTCAAACAGGAGGTC 740
|||||
743 GlnPhe 744
|||||
741 AGGTGG 746

```

seq_name: gb_est98:BG258462

seq_documentation_block:

LOCUS BG258462 777 bp mRNA EST 13-FEB-2001
DEFINITION 602379941F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4510378 5',
mRNA sequence.

ACCESSION BG258462

VERSION BG258462.1 GI:12768191

KEYWORDS EST

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 777)

AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10392 row: b column: 11

High quality sequence start: 4

High quality sequence stop: 678.

FEATURES

source

1..777

/organism="Homo sapiens"

JOURNAL
COMMENT

Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
[RessourcenzentrumPrimatenbank, Berlin, Germany](http://RessourcenzentrumPrimatenbank.Berlin.Germany) (web address:

FEATURES

source

```

alignment_scores:
  Quality: 770.00      Length: 190
  Ratio: 4.208         Gaps: 0
  Percent Similarity: 96.316  Percent Identity: 78.421

alignment_block:
  US-09-512-581-2 x AI794456 ..

Align seg 1/1 to: AI794456 from: 1 to: 570

712 ProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLysTy 728
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2 CCAGTACTGCAGCACAAAGCAAAAGGGCCCTCTCGCAGGCGCAAGTA 51

728 rAlalleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheA 745
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
52 TGCCATCCAGTGGGTCCACGCCCATGTTTCCACAGAGACACACTTGT 101

745 laGlnIlePheGluProLeuHisLysSerLysAspProSerAsnLeuGlu 761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 CACAGATATTTGACCTCTGCATTAAGGCTTAGACACACAGACATGGAG 151

762 HisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPr 778
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 CAGCTCATCAGCGCTTTAAACAGCACTGGGACACTGGCCATGCTGGCCCC 201

```

Align seg 1/1 to: AI666974 from: 1 to: 580

```

739 SerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24 AACAGAGACACACACATTTGCACAGATATTTGAGCTCTGCATAAGGGTCT 73
755 uasProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyH 772
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
74 AGACACAGACACACATCGAGCAGCTCATCAGCGCTTTAAGACACATGGGAC 123
772 isLeAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrp 788
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
124 ACTGGCCATGCTGGCCCCAGACAGATTTGCTGCTCCGCTCAAACTTTTA 173
789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGl 805
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 GTGGCCAAATTCATCGTTAAAGACCTGCTAATGAAGACCGGATTCCTGG 223
805 yLysLysThrThrLysLysLeuTrpValProAspGluGluValSerProGlu 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 CAAGAAACACCAAGCTGTGGTGCCAGATGACGAAGTCTCTCCTCGAAA 273
822 hMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGly 838
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 CATCACCAAGATTCAGGCTCTGAACCTGATGATGCTGGCTGCTCGGT 323
839 MetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuTh 855
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 GTGAACAACAACAGAGTAAATCTGGAACCTCAACTCTCGGATGTTGAC 373
855 rThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerL 872
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 GGCAAATCCTGAGCAGCGATGGAGATCTGCGGAGCAGGCAAGATGGGA 423
872 ysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLys 888
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 AGCGGACATGCTCTGCTGCGTCTGGCAGCGCTGCGATTCCTCGT 473
889 LeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGl 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 CTCGCTCAGAGCGCTGCTACCATGATGATCATCACCTGGAGCAGTACCA 523
905 nLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValP 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
524 GCCTCGGCCCTCGTCTCATCAACGACGAGTGCTACCAGGTGCGACAGGCTN 573
922 heAla 923
|||
574 TCGCT 578

seq_name: gb_est25:AI794456

seq_documentation_block:
LOCUS AI794456 570 bp mRNA EST
DEFINITION fc44f03.y1 zebrafish WashU MPMG EST Danio rerio cdi
sequence.
ACCESSION AI794456
VERSION AI794456.1 GI:5342172
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Cypriniformes; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 570)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F.,
S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyl
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.
and Wilson, R.
WashU zebrafish EST Project 1998

```


/tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker: Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 169 a 101 c 95 g 105 t
 ORIGIN
 alignment_scores:
 Quality: 786.00 Length: 156
 Ratio: 5.071 Gaps: 0
 Percent Similarity: 99.359 Percent Identity: 98.718

alignment_block:

US-09-512-581-2 x AA282190

Align seg 1/1 to: AA282190 from: 1 to: 470

1075 llelleMetSerLysSerThrTyrSerLeuGluSerProLysAspPr 1091
 1 ATCATCATGTCAGAGAGTACTACATACAGTTTGGAAATCTCTAAAGACCC 50

1091 oValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnT 1108
 51 GGTACTACAGCTCGTTCTTCTACCTACACCTGACAGAAATTCAGTAACA 100

1108 hrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124
 101 CCAAAATATCTGCCTCTCTGAAATGAATCATATTTTCTACTCTCTGGAANA 150

1125 ProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAl 1141
 151 CCTAAACACCAATGTTCTAGGAGCTGTAAACAGCCACTTTCATCAGC 200

1141 agLysLysGlnSerGlnThrLysSerArgMetGluThrValSerAsnA 1158
 201 AGGCAAGCAATCTCAGACCAATCATCAGATGGAACCTGTAACCAATG 250

1158 laSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeu 1174
 251 CAAGCAGCAGCTCAATCCAGCTCTCTGGAAGAAATAAAGGGAGGCTT 300

1175 AspSerSerGluMetAspHisSerSerGluAsnGluAspTyrThrMetSerSe 1191
 301 GATAGTCTGAAATGGATGCAGTCAAGTGAATGAAGATTACACAAATGCTCTC 350

1191 rProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValA 1208
 351 ACCTTTGCGGGGAAAAAAGTGACAGAGAGACGACTCTGATCTGTGTA 400

1208 rgSerGluLeuGluLysProArgGlyArgLysLysThrProValThrGlu 1224
 401 GGTCTGAATGGAGAGCCTAGAGCAGGAGAAAAAACGCTCGTCACAGAA 450

1225 GlnGluGluLysLeuGly 1230
 451 CAGGAGGAGAAATAGGT 468

seq_name: gb_est23:AI666974

seq_documentation_block:

LOCUS AI666974 580 bp mRNA EST 18-MAY-1999

DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

FEATURES
 source

BASE COUNT 145 a 166 c 151 g 117 t 1 others
 ORIGIN

alignment_scores:
 Quality: 783.00 Length: 185
 Ratio: 4.399 Gaps: 0
 Percent Similarity: 96.216 Percent Identity: 79.459

alignment_block:
 US-09-512-581-2 x AI666974

fc24a12.y1 Zebrafish Washu MPIMG EST Danio rerio cDNA 5', mRNA
 sequence.
 AI666974
 AI666974.1 GI:4805330
 EST.
 zebrafish.
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Rasbora; Danio.
 1 (bases 1 to 580)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Bddy
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.
 Washu Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@wustl.wustl.edu
 CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 480.
 Location/Qualifiers
 1..580
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="Zebrafish Washu MPIMG EST"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRF"
 /note="Vector: pSPORT1; Site.1: NotI; Site.2: SalI; 1st
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 [5'-PGACTAGTCTAGATCCGAGCGCCGCTTTTCTTTT3'];
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

751 Leu.HisLysSerLeuAspProSerAsnLeuGluHisLeuLeuThrProL 767
 645 CTGTCAGGCTAGCTCTGAACGAGGAGGTCAGAGAGAGGTCATCATCC 596
 767 euValThrLeuGluHisLeuAlaLeuLeuAlaProAspGlnPheAla 783
 595 TCGTGTCTACTCGCCACATCTTCATCTGCGCCAGATCAGTTGCTTCA 546
 784 ProTrpLysSerTrpValAlaThrPheLeuValLysAspLeuMetAs 800
 545 CGGATGAAGTCCATTGTGGTAACATTCATCGTGAAGGACCTGCTCAT 496
 800 nAspArgLeuProGlyLysLysThrLysLeuTrpValProAspGlu 817
 495 TGACAGATCGGTGGGAAACAAGATGGGAAGCTGTGGACCACTGATG 446
 817 luValSerProGluThrMetValLysLeuGlnAlaLeuLysMetVal 833
 445 AAGTCTACCTGAAGTTCTCGCTAAGTGCAGGCCATCAAGCTGTGGT 396
 834 ArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 850
 395 CGTGTGTTAGGAATGAAGAACCAACCAATCAAGTCGGCCAACTCCAC 346
 850 leuArgLeuLeuThrThrLeuHisSerAspGlyAspLeuThrGlu 867
 345 CCTCGCTGCTGTTCTGGGCATCTTGGTCAGCGGAGGAGACCTCAGAG 296
 867 InGlyLysLeuSerLysProAspMetSerArgLeuArgLeuAlaGly 883
 295 AGAAGATCATGTAATCAGACATCTCTCGCTGAGGCTGCGCCAGGT 246
 884 SerAlaLeuValLysLeuAlaGlnGluProCysTyrHisGluLeuLeu 900
 245 GGAGCCATTATGAAGTTAGCCAGGAGCTGTGTACCATGACATCATC 196
 900 leuGluGlnTyrGlnLeuCysAlaLeuAlaLeuAsnAspGluCysTyr 917
 195 CCTGAACAGTTTTCAGCTCTCGGCTTGTATCAATGACGAGTCTACC 146
 917 InValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArg 933
 145 AGTTTCGTCAGATCTTTGCTCAGAGCTGCACCTGGCTCTGTCAAGCT 96
 934 ArgLeuProLeuGluTyrMetAlaLeuCysAlaLeuCysAlaLysAsp 950
 95 GTGCTTCCCTCGAGTACTTGGCGGTGTGGCTGTGTGGTGGCAAGGACC 46
 950 oValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
 45 GGTGAAGGAGCGCGCGCCCGCCGCGCCGACAGTGCCTC 9

seq_name: gb_est102:BG571128

seq_documentation_block:
 LOCUS BG571128 888 bp mRNA EST 10-APR-2001
 DEFINITION 602591748F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4714234 5',
 mRNA sequence.
 ACCESSION BG571128
 VERSION BG571128.1 GI:13578781
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 888)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLML at:
 http://image.llnl.gov

Plate: LLCMI558 row: p column: 11
 High quality sequence stop: 688.

FEATURES

Location/Qualifiers

source

1..888

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4714234"

/lab_host="NIH_MGC_79"

/note="Organ: placenta; Vector: pDNR-LIB (Clontech);

Site_1: SfII (ggccgctcgcc); Site_2: SfII (ggccatagggc

): 5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-ATTCTAGAGCGGCGCCGACATG-dT(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH_MGC

Library."

BASE COUNT 347 a 169 c 223 g 149 t

ORIGIN

alignment_scores:

Quality: 806.00 Length: 160

Ratio: 5.037 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-512-581-2 x BG571128

Align seg 1/1 to: BG571128 from: 1 to: 888

1232 AspAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnAr 1248

2 GATGACTTGACTAAGTTGGTACAGCAACAGAACCTAAAGGAGTCAGCG 51

1248 gSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnTrpP 1265

52 AAGTCGAAAGAGGCGCATACGCTTCAGAAATCTGATGAACAGCAGTGC 101

1265 roGluGluLysArgLeuLysGluAspLeuGluAsnGluAspGluGln 1281

102 CTGAGGAAAGAGGCTCAAGAGAGATATATTAGAAATGAAGATGAACAG 151

1282 AsnSerProProLysGlyLysArgGlyArgProProLysProLeuGl 1298

152 AATAGTCCGCCAAA.AAGGTAAGAGAGCGCCGCCACCAACCTCTTGG 200

1298 yGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlys 1315

201 TGGAGGTACACCAAAAGAGAGCCCAACATGAAACTTCTAAAAAGGAA 250

1315 erLysLysLysSerGlyProProAlaProGluGluGluGluGlu 1331

251 GCAAAAAAATCTGGACCTCCAGCAGGAGGAGGAGGAGGAGGAGGAA 300

1332 ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgVa 1348

301 AGACAAAGTGGAATACGGAACAGAGTCCAAAGCAACACACCCAGCT 350

1348 lSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleG 1365

351 GTCAAGGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 400

1365 luSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThr 1381

www-bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA-Yes.

FEATURES

source Location/Qualifiers

1..490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3088558"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub8 is a subtracted library derived from NCI_CGAP_Sub5. The NCI_CGAP_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI_CGAP_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE clone ids 2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI_CGAP_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI_CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_LIB="NCI_CGAP_Lu5
TAG_TISSUE="lung
TAG_SEQ="CAAC"
167 a 96 g 148 t
TAG_79 c

BASE COUNT
ORIGIN

alignment_scores:

Quality: 816.00 Length: 159
Ratio: 5.197 Gaps: 0
Percent Similarity: 98.742 Percent Identity: 98.742

alignment_block:

US-09-512-581-2 x BF511856 ..

Align seg 1/1 to: BF511856 from: 1 to: 490

894 CysTyrHisGluLeuLeuGlnTyrGlnLeuCysAlaLeuAl 910
|||||
13 TGTTACCATGAATATCATCATTAGAACATATCATGCTATGTCATAGC 62
910 alleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLeuH 927
|||||
63 TATCAACGATGAATGTATCAAGTAGACAAAGTGTGGCCAGAACTTC 112
927 isLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCys 943
|||||
113 ACAAGGCGCTTTCCCGTTAGCGCTTCCACTTGAGTATATGCGCAATCTGT 162
944 AlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgG1 960
|||||
163 GCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCATGCTAGGCA 212
960 nCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 977
|||||
213 ATGTTTGGTGAATAATATTAATGTAGGGGGGAGTATCTGAACGACATG 262
977 laAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993
|||||
263 CAGCTGTAGTGAATAATATTGCTCTCTACACGAGATGTTGTTC 312
994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
|||||
313 TATACAAATTCACCTTTTGGCACATGATGATGTCACCAAGTACAGGA 362

1010 pileGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluI 1027
|||||
363 TATTGAACACCTTAAGATGTTAAAGAAATGCTTTGGTTTCTGGAAA 412
1027 leLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysMet 1043
|||||
413 TATTAAATGGCTAAATAAGAAATAACAGTCACGCTGTATCAGAAAGATG 462
1044 valGluAsnIleLysGlnThrLysAsp 1052
|||||
463 GTAGAAAATATTAAACAAACAAAAGAT 489
seq_name: gb_est41:AW013141

seq_documentation_block:

LOCUS AW013141 746 bp mRNA EST 10-SEP-1999
DEFINITION SRN011SK Winter flounder stomach Pleuronectes americanus cDNA clone sequence.
ACCESSION AW013141 GI:5861919
VERSION AW013141.1
KEYWORDS EST.
SOURCE Pleuronectes americanus.
ORGANISM Pleuronectes americanus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes; Pleuronectoidi; Pleuronectidae; Pleuronectes.
REFERENCE 1 (bases 1 to 746)
AUTHORS Douglas,S.E., Gallant,J.W., Bullerwell,C.E., Wolff,C., Munholland,J. and Reith,M.E.
TITLE Winter flounder expressed sequence tags: Establishment of an EST database and identification of novel fish genes
JOURNAL Marine Biotechnology (1999) In press
COMMENT Contact: Reith M
Marine Biology
NRC Institute for Marine Biosciences
1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada
Tel: (902) 426-8276
Fax: (902) 426-9413
Email: michael.reith@nrc.ca
Seq primer: M13 Forward.

FEATURES

source Location/Qualifiers
1..746
/organism="Pleuronectes americanus"
/db_xref="taxon:8263"
/clone="SRN011SK"
/clone_lib="Winter flounder stomach"
/sex="female"
/dev_stage="adult"
/note="Organ: stomach"
BASE COUNT 171 a 192 c 209 g 174 t
ORIGIN

alignment_scores:

Quality: 806.00 Length: 247
Ratio: 3.784 Gaps: 3
Percent Similarity: 86.235 Percent Identity: 67.611

alignment_block:

US-09-512-581-2 x AW013141/rev ..

Align seg 1/1 to reverse of: AW013141 from: 1 to: 746

720 LysGlyProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaI1 736
|||||
745 CAAAGCCCCAAAGCCAGGGGCTTTCCCAACTGGTAATTCCTCCATGCCCA 696
736 ePheSerSerLysGlu.....ThrGlnPheAlaGlnIle.PheGluPro 750
|||||
695 TTTTAAAAACAAGGAAGTGCAAGCTGGCAGACATTTTGTAGCTT 646


```

|||||
603 TGTGTTTAACTAGTCAAGTGTGTGGCAAGGCA..... 640
1085 LeuGluSerProLys... AspProValLeuProAlaAArgPhePhe ThrG 1100
:::|||||::: ||| ||| :::||||| |||
641 ...GACTCACCAAGGGAACCAAGTCTCCCAATCAAGTCTTTAACAAC 687
1100 InProAspLysAsnPhe.SerAsnThrLysAsnThrLysProGluMe 1116
:::||| ||| ::::: |||::: |||
688 CGGCTGAGAGAGCCTTCTGTTAGGACAGAGCGCTATATTTAGGAAACAA 737
1116 tlysser PhePheThrProGlyLysProLysThrThrAsnVal 1130
:::||| |||::: |||::: ||| |||::: |||
738 CAACAGTGTCTGTTGAACGAGAGACCAACGCCCTACGGGGGTA 781

```

seq_name: gb_est76:BE551003

```

seq_documentation_block: 535 bp mRNA EST 10-AUG-2000
LOCUS BE551003 7667b03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:323261 3',
mRNA sequence.
ACCESSION BE551003
VERSION BE551003.1 GI:9792695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 535)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 475.
Location/Qualifiers
1. 535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:323261"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP-Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
141920-141791 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

```

BASE COUNT 183 a 90 c 108 g 153 t 1 others
ORIGIN

```

FEATURES

source

```

1. 535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:323261"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP-Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
141920-141791 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

alignment_scores:

```

Quality: 817.00 Length: 170
Ratio: 4.922 Gaps: 0
Percent Similarity: 97.647 Percent Identity: 95.294

```

alignment_block:

US-09-512-581-2 x BE551003

Align seg 1/1 to: BE551003 from: 1 to: 535

```

905 GlnLeuCysAlaLeuAlaLeuAsnAspGluCysTyrGlnValArgGlnVa 921
|||||
27 CAGCTATGTGCATTAGCTATCAACGATGATGCTNATCAAGTAGACAAGT 76
|||||
921 lPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuG 938
|||||
77 GTTGGCCCAAGAACTTAC.AAAGGCGCTTTCCCGTTTACGGCTTCCATT 125
|||||
938 lutyMetAlaLeuCysAlaLeuCysAlaLysAspProValLysGluArg 954
|||||
126 AGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTTAAGGAGA 175
|||||
955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArg 971
|||||
176 AGAGCTCATGCTAGGCAATGTTGGTGAATAATATAAATGTAAGCGGGA 225
|||||
971 uTrpLeuLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuP 988
|||||
226 GTATCTGAAGCAGCATGCAGCTGTAGTGAATAATATTGTCTCTCTAC 275
|||||
988 roGluTyrValValProTyrThrIleHisLeuLeuAlaHisaspProasp 1004
|||||
276 CAGAGTATGTTCTCATATACATTCACCTTTGGCAGCATGCCAGAT 325
|||||
1005 TyrValLysValGlnAspIleGlnLeuLysAspValLysGluCysLe 1021
|||||
326 TATGTCAAAGTACAGATATTAACAACCTTAAGATGTTAAGAATGTCT 375
|||||
1021 uTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHis 1038
|||||
376 TTGGTTTGTCTGGAATATTATGCTGAAATGAAATAACAGTCACG 425
|||||
1038 laphileargLysMetValGluAsnIleLysGlnThrLysaspAlaGln 1054
|||||
426 CTTTATCAGAAAGATGGTCGAAAATTTTAAACAAACAAAGATGCCAA 475
|||||
1055 GlyProAspAlaLysMetAsnGluLysLeuTyrThrValCysaspVa 1071
|||||
476 GCACCATGATGCAAAATGAATGAAAACTGTACACTGTGTGTGTCT 525
|||||
1071 lAlaMetAsn 1074
|||||
526 TGCATGAAT 535

```

seq_name: gb_est88:BF511856

```

seq_documentation_block: 490 bp mRNA EST 06-DEC-2000
LOCUS BF511856 UI-H-B14-aps-d-12-0-UI.sl NCI_CGAP_Sub8 Homo sapiens cDNA clone
DEFINITION IMAGE:3088558 3', mRNA sequence.
ACCESSION BF511856
VERSION BF511856.1 GI:11595154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 490)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:

```



```
1154 ThrValSerAsnAlaSerSerSerAsnProSerSerProGlyArgII 1170
|||||
480 ACTGTAAAGCAATCAAGCAGCAGCAGCTCAATCCAAAGCTCTCTCGGAAGAT 431
|||||
1170 eLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspT 1187
|||||
430 AAAGGGAGGCTTGATGTTCTCAATGGATCACAGTGAATGAAGATT 381
|||||
1187 yThrMetSerSerProLeuProGlyLysSerSerAspLysArgAspAsp 1203
|||||
380 ACACAATGCTTCACCTTTCGCGGGGAAAAAAGTGACAAAGAGAGACGAC 331
|||||
1204 SerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysTh 1220
|||||
330 TCATGCTTGTAGGCTCTGAATTTGGAGAACCCCTAGAGCGAGGAAAAAAC 281
|||||
1220 rProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysL 1237
|||||
280 GCCCGTCACAGACAGGAGGAGAAATTAGGTATGGATGACTTCGACTAAGT 231
|||||
1237 euValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGly 1253
|||||
230 TGTTACAGGAACAGAAACCTTAAGGAGCTCAGCGAAGTCGGAAGAGGC 181
|||||
1254 HisThrAlaSerGluSerAspGluGlnGlnTrpProGluLysArgLe 1270
|||||
180 CATACGGCTTCAGAACTGTATGAAACAGCAGTGGCCTGAGGAAAGAGGCT 131
|||||
1270 uLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysL 1287
|||||
130 CAAAGAAGATATATTAGAAATCAAGATGAACAGATAGTCCGCCAAAA 81
|||||
1287 ySGLysArgGlyArgProProLysProLeuGlyGlyThrProLys 1303
|||||
80 AGGGTAAAGAGCGCCAGCACCAACCAACCTCTTGGTGGAGGTACACCAAAA 31
|||||
1304 GluGluProThrMetLysThrSerLysLys 1313
|||||
30 GAAGAGCCAAAAATGAACACTTCTAAAAA 1
```

seq_name: gb_est97:BG176414

```
seq documentation block:
LOCUS BG176414 910 bp mRNA EST 06-FEB-2001
DEFINITION 602338082F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4460901 5',
```

mRNA sequence.

ACCESSION BG176414

VERSION BG176414.1 GI:12683026

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10263 row: d column: 22

High quality sequence stop: 628.

Location/Qualifiers

1..910

/organism="Mus musculus"

/strain="FVB/N"

FEATURES

source

```
/db_xref="taxon:10090"
/clone="IMAGE:4460901"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 276 a 199 c 225 g 210 t
ORIGIN
```

alignment_scores:

Quality: 827.50 Length: 268

Ratio: 3.662 Gaps: 8

Percent Similarity: 84.328 Percent Identity: 63.433

alignment_block:

US-09-512-581-2 x BG176414 ..

Align seg 1/1 to: BG176414 from: 1 to: 910

```
874 AspMetSerArgLeu ArgLeuAlaAlaGlySerAlaIleValLysLeuA 890
|||||
3 GATATGCTCTGCTGGGATTAGCTGCTGCTAGTGCCTATTAAGCTTG 52
|||||
890 lagGlnGluProCysTyrHisGluIleThrLeuGluGlnTyrGlnLeu 906
|||||
53 CTCAGGAACCTTGCTACCATGAAATATTACCACCAGAACAGTTTCAGCTC 102
|||||
907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAl 923
|||||
103 TGTGCACCTGGTTATTATGATGAGTGTCTACCAAGTAGGAGCAGATATTGC 152
|||||
923 aGlnLysLeuHisLysGlyLeuSer ArgLeuArgLeuProLeuGluTyr 939
|||||
153 CCAGAACCTTCATAAGGCACCTTGGTGAAGTTGCTCTCCACCTGGAGTAT 202
|||||
940 MetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAl 956
|||||
203 ATGCCATCTTTGCTTTGTGTCGCAAGACCCCTGTGAAGGAAAGGAGAGC 252
|||||
956 aHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrL 973
|||||
253 ACACGCTCGGCAGTGTGTGTTAAAGAACATCAGCATCCGCGAGGAGTACA 302
|||||
973 euLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGlu 989
|||||
303 TCAACAGAACCCCATGCGCCCATGAGAAATATTGTCACTGCTGCTCGTAA 352
|||||
990 TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa 1006
|||||
353 TATGTGGTTCATACATGATGATTCACCTCTCTAGCCCATCATCTGATTTTCAC 402
|||||
1006 LysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp 1023
|||||
403 ACGGTCAACAAGATGTTGATCAACTTCGTATATAAAAGAGTGCCTGTGGT 452
|||||
1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPhe 1039
|||||
453 TTTATGTTGAAGTCTTAATGACAAAGAAATGAAACACACAGCCATGCATTC 502
|||||
1040 IleArgLysMetValGluAlaAsnIleLysGlnThrLysAspAlaGlnGlyPr 1056
|||||
503 ATGAAAAAGATGCGAGAGATATCAAGCTAACACAGAGACGCCAGTCTCC 552
|||||
1056 oAspAspAlaLysMetAsnGluLysLeuTyr.....ThrV 1068
|||||
553 CGATGAATCCAAAGCAATGAAAACTTATACGTTTCTGACGTGGCTCTG 602
|||||
1068 alCysaspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
```


Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 58 row: L column: 16

Seq primer: ATTATGCTGACACTATAG.

FEATURES

Location/Qualifiers

source

1..560

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 2BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."

193 a 96 c 121 g 150 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 830.00

Length: 186

Ratio: 4.637

Gaps: 0

Percent Similarity: 96.237

Percent Identity: 82.258

alignment_block:

US-09-512-581-2 x BE756042 ..

Align seg 1/1 to: BE756042 from: 1 to: 560

334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350

1 GTGAAATTTGCCAGTCACCTGTTTAAATGAATCAATCCAGATTTAGCAAAAGGA 50

350 pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleA 367

51 TCTCACAGATATTTCGAAAGTTCGATCCCATGATCCAGAGAGAGCTATTTC 100

367 rGHisAspValIleValSerIleValThrAlaAlaLysAspIleLeu 383

101 GTCATGATCTTATTGTTACTATAATAAATCTGCTGCCAAAAGACACCTTGC 150

384 LeuValAsnAspHisLeuAsnPheValArgGluArgThrLeuAspIly 400

151 TTAGTGAATGATCAGCTACTTGGCTTTGTAGAGAAAGAACACTGGACAA 200

400 sArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrL 417

201 ACGTGGCGAGTAAAGAAAGAGCTATGATGGTCTGGCTCAGCTTTATA 250

417 ySLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaLysGlnIle 433

251 AGAAATACTGCTTCTATGCTGAAGCAGGAAAGAGCTGCAGAGAAAGTC 300

434 AlaTrpIleLysAspIleLeuHisIleTyrTyrGlnAsnSerIleAs 450

301 ACCTGATAAGGACAACTTTTGCATATATATATCAAAATAGCATCGA 350

450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467

351 TGACAAACTATTGTTAGAGAAATCTTCTCAATATCTTGTGCCCCACA 400

467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483

401 ACCTGGAACAGAGAGAGAAATGAAATGCTTGTATTATTATATGCTACT 450

484 LeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAs 500

451 TTGGATCCAAATGCTGTCAAAGCTCTCAATGAATGTGGAAGTGCAGAA 500

500 nLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProL 517

501 CATGCTTAGAAGTATGATGACGAGAACTGTTGGATTGACACAGCAGCTTA 550

517 ySThrAsp 519

551 CATCAGAA 558

seq_name: gb_est23:AI655429

seq_documentation_block:

LOCUS AI655429 480 bp mRNA EST 14-DEC-1999

DEFINITION ts98f06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2239331 3',

mRNA sequence.

ACCESSION AI655429

VERSION AI655429.1 GI:4739408

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (Bases 1 to 480)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1048 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 455.

Location/Qualifiers

1..480

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2239331"

/clone_lib="NCI_CGAP_GC6"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; plasmid DNA from the normalized library

NCI_CGAP_GC4 was prepared, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from a pool of 5,000 clones made

from the same library (cloneids 1257096-1258631,

1469064-1470983, and 1475592-1476743). Subtraction by

Bento Soares and M. Fatima Bonaldo."

Bento Soares 82 a 123 c 84 g 191 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 828.00

Length: 160

Ratio: 5.208

Gaps: 0

Percent Similarity: 99.375

Percent Identity: 99.375

alignment_block:

US-09-512-581-2 x AI655429/rev ..

Align seg 1/1 to reverse of: AI655429 from: 1 to: 480


```

seq_name: gb_est102:BG574948

seq_documentation_block:
LOCUS BG574948 886 bp mRNA EST 10-APR-2001
DEFINITION 602598035f1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4706697 5',
mRNA sequence.
ACCESSION BG574948
VERSION BG574948.1 GI:13582601
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 886)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10575 row: f column: 10
High quality sequence stop: 637.
FEATURES
location/Qualifiers
1..886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4706697"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 282 a 167 c 221 g 216 t
ORIGIN

alignment_scores:
Quality: 850.00 Length: 279
Ratio: 3.542 Gaps: 6
Percent Similarity: 86.022 Percent Identity: 65.233

alignment_block:
US-09-512-581-2 x BG574948 ..
Align seg 1/1 to: BG574948 from: 1 to: 886
400 LysArgTIPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTy 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 AAACGGTGGCGATAGAAAAGAGCTATGATGGTCTGGCTCAGCTTTA 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 rLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnI 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 TAAGAAATACTGCTTCATGGTGAAGCAGGAGGAAGCTGCAGAGAAAG 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 leaLatrPileLysAspLysLeuHisIleTyrTyrGlnAsnSerIle 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
129 TCAGCTGGATAAGGACAACTTCGCATATTATTATTCAGACAGCAT 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 AspAspArgLeuValGluArgIlePheAlaGlnTyrMetValProHi 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 GACGACAACTGTTGGTAGAGAAACTTTGCTCAGTATCTTGTCCCCA 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 asnLeuGluThrGluArgMetLysCysLeuTyrTyrLeuTyrAlat 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

229 CAACCTGGAAACACAGAGAGAAATGAATGCTTATATTTACTTATGCTA 278
483 hrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGln 499
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 GTTTGGATCAAAATGCTGTAAAGCTCTCAACGAATGTGGAAAGTGCAG 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 AsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnPr 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
329 AACATGCTTCGGAGCCATGTACCGGAACATTGATTTGGCACAAGCACC 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 oLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIle 533
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
379 TACATCAGAGGCTAACTGTCTGCCATGTTTGGAAAACATGATCACCAT 428
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
533 hrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 549
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 CAAAGAATTTGCTGACCCCGGGAAGACACAAAGATTTGTGAAGAAAT 478
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
550 ThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluVal 566
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 AACAGGTTCTCGCGCATGATGAGAACTTCGCTCAGTTGGAGTTATT 528
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
566 uValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGlu 582
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 AATTAGCCCAACGTCTTCTTCAACAAGCAGATATTGGGTGAGAGAA 578
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 IleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeu 599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
579 ATAGCCCGGAACTTGCAATCTTAAGGAACAAAAATCCCTT.TTTCFAGA 627
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
599 uMetIleLysPheLeuLeuGlu..ArgIleAlaProValHisIleAspTh 615
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
628 GATGGTCAATTTCTGTGGAACAGACATCGGACTGTGCCCTTGATTC 677
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
615 rGluSerIleSerAlaLeuLysGlnValAsnLysSerIleAsp..G 631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
678 AGAAGCCATAAGTGCACCTAGTAGAACATTCGCTGAATTAAGTCATAAG 727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 lyThrAlaAspAsp...GluAspGluGlyValProThrAsp.GlnAla 646
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
728 GACACCCAGATGATGACGAGGAGGAGGTGTAAAGTCAGAAATACAGGTAT 777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
646 eArgAla.GlyLeuGluLeuLysValLeuSerPheThrHisProIle 662
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
778 CCGTCAAGGACTGAACCTCTTAAGGCTTGGGCTTTACACATTCCAC 827
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 SerPheHisSerAlaGluThrPheGluSer 672
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828 CTGGTTCAATCTGGAGAGACCTATGAGTCC 857
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est78:BE756042
seq_documentation_block:
LOCUS BE756042 560 bp mRNA EST 15-SEP-2000
DEFINITION 210069 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE756042
VERSION BE756042.1 GI:10170034
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 560)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.W. and Keeler,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

```



```

|||||
2  TTTTATTACTTGAGAACATTTGCTGGGCAAGGCATATAGCATATGCTT 51
145 eGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuP 162
52  TGAGTTAGAGATAGCAATGAATTTTATCCAGCTATACAGAACCTTAT 101
162 heSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetVal 178
102 TTTTCAGTTATAACATGGCCACAATCAGAAAGTCCATATGCACATGGTA 151
179 AspLeuMetSerSerIleLeuGluGluGluGluGluGluGluGluGlu 195
152 GACCTTATGAGCTCTATTATTCTGAAGGTGATACAGTGTCTCAGGAGCT 201
195 uLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnL 212
202 TTTGGATACGGATTTTGTAAATCTGTACCTGCTCATAGAAATTTAAGCA 251
212 yGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAla 228
252 AGCAGCATATGATTTGGCAAGGCTTTACTGAAGGACAGCTCAAGCT 301
229 IleGluProTyrIleThrPheAsnGlnValLeuMetLeuGly 245
302 ATTGAGCCATATATTACCAATTTTATATCAGGATCTGTGCTTGGAA 351
245 sThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeuGluL 262
352 AACATCTATCAGCGATTTGTGACAGCATGTCTTTGACTTAATTTTGGAGC 401
262 euTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
402 TCTACAAATATTGATGATCATTTGCTGCTCTCTGCTTTACCCAGCTTGAA 451
279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValLysLe 295
452 TTTAAACTAAAGAGCAATGATATAGGAGCGCTACAAGCTGATAAACT 501
295 uLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnL 312
502 ACTGGC....ATGTTAGTGGCAAGGATTCAGAAAGTGGCTTCT..... 540
312 ysProLeuTyrGlnCysTyrLeuGlyArgPheAsnAspIleHisValPro 328
541 .....CCA 543
329 IleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisPr 345
544 ATCCGCTGGAATGTGTG...TTCGCTAGCATTCCTCATGAACCATCC 590
345 OAspLeuAlaLys 349
591 TGATTTAGCACAA 603

```

seq_name: gb_est100:BG391029

```

seq_documentation_block: 885 bp mRNA EST 12-MAR-2001
LOCUS BG391029
DEFINITION 602417722F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4537108 5',
mRNA sequence.
ACCESSION BG391029
VERSION BG391029.1 GI:13284477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10461 row: 1 column: 05
High quality sequence stop: 695.

FEATURES

source
1. 885
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4537108"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 277 a 167 c 202 g 239 t
ORIGIN

alignment_scores:

Quality: 884.00 Length: 259
Ratio: 3.794 Gaps: 3
Percent Similarity: 89.961 Percent Identity: 72.973

alignment_block:

US-09-512-581-2 x BG391029 ..

Align seg 1/1 to: BG391029 from: 1 to: 885

```

761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAl 777
|||||
6  GACAACTTATAACTCCATTTAGTTTC.ATGGCCACATTTCTATGTAGC 54
777 aProAspGlnPheAlaAlaProTyrLysSerTrpValAlaThrPheIleV 794
|||||
55 ACCAGATCAGTTTGTCTCCCAATGAAATCTGTAGTAGCAAAATTTATTG 104
794 aLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrLys 810
|||||
105 TGAAGATCTGCTTAATGAATGACAGGTCAACAGGTGAAAGAAATGCAAAA 154
811 LeuTrpValProAspGluGluValSerProGluThrMetValLysIleG 827
|||||
155 CTGTGCTCTCCAGATGAAGAGTTTCCCTGGAAGTACTAGCAAGGTACA 204
827 nAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHis 844
|||||
205 GGCAATTAACCTTCTGTAGGTGGCTGTG. GGTATGAAAAACAACCACT 253
844 erLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
|||||
254 CTAATCTGCAATTAACCCCTTCGGTTATTATCAGCGATGTGGTTAGT 303
861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerAr 877
|||||
304 GAGGGTACCTGACAGAGCAAAAGAGATCAGTAATCTGATATGCTCG 353
877 gLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluPro 894
|||||
354 CTTGCGATTAGTCTGTTAGTGCATTAATGAAGCTTGTCTCAGGAACCTT 403
894 yTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAla 910
|||||
404 GTTACCATGAATATTATACCCCAAGAGTTTCAGCTCTGTGCACTGTGT 453

```



```

:::|||||
602 ACCAGGTTCTCGCGATGATGAGAACTTCGGTCTCAGTTGGAGTTATTA 651
567 ValSerProThrCysSerCysLysGlnAlaGluClyCysValargGlu 582
652 ATTAGCCCAACCTGTACTTCAACAACAGCAGATATCGGTGTGAGAGAA 699

```

seq_name: gb_est92:BF789050

```

seq_documentation_block:
LOCUS BF789050 902 bp mRNA EST 12-JAN-2001
DEFINITION 602104907F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223040
5', mRNA sequence.
ACCESSION BF789050
VERSION BF789050.1 GI:12094086
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
1 (bases 1 to 902)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9810 row: n column: 01
High quality sequence stop: 655.
Location/Qualifiers
1..902
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:4223040"
/lab_host="NCI_CGAP_Kid14"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

```

```

FEATURES
Source
1..902
274 a 156 c 233 g 239 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
Quality: 899.00 Length: 262
Ratio: 3.777 Gaps: 5
Percent Similarity: 90.840 Percent Identity: 72.901

```

alignment_block:

US-09-512-581-2 x BF789050 ..

Align seg 1/1 to: BF789050 from: 1 to: 902

```

208 LysAsnLeuAsnLysGlnAlaTyAspLeuAlaLysAlaLeuLysAr 224
|||||
2 AAGAACCTTAATTAACAGTCTTCGACCTTGCAAAAGCTCTACTGAAAG 51
224 gThAlaGlnAlaIleGluPro..TyrlThrThrPhePheAsnGlnVa 240
|||||
52 GACAGTCCAGACATTGAAGCATGTAT...GCCAATTTTTCATCAAGT 98
240 lLeuMetLeuGlyLysThrSerLeSerAspLeuSerGluHisValPhe 257
|||||
99 CCTGGTCTGGGAGATCATCAGTCAGCAGCCTGTCTGAAACAGCATTTG 148

```

```

257 spLeuIleLeuGluLeuTyAsnIleAspSerHisLeuLeuLeuSerVal 273
|||||
149 ATCTGATTCAGGAACCTTTTCTCT.ATCGATCCTCAGTTACTGTTATCTGTC 197
274 LeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLe 290
:::|||||
198 ATGCCACAGCTTGAATTCAAACTGAAGCAACGATGGTGAAGAACGCT 247
290 uGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluL 307
|||||
248 AGCTGTGGTTCGACTCTCGCAAAATTTGTCGGCTCTAAAGATTCAGATT 297
307 euAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGlyArgPheAsn 323
|||||
298 TAGCACACAGATCGGCTCTCTGGCAGTCTTCTTGGCGGATTTAAT 347
324 AspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCy 340
|||||
348 GACATTCATCTCTCTGTGAGTTAGAAAGTGTGAAGTTGCCAGCCACTG 397
340 sLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysV 357
|||||
398 TTTGATGAATCAATCCAGACTTAGCGAAGGATCTGACAGAAATATTGAAG 447
357 alArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 373
|||||
448 TTAGGTTCATGATCCAGAGAGCCATTCTCATGATGTTATTGTTACT 497
374 Ile.ValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuL 390
|||:::|||||
498 ATAATAACAGCTGCCAAAAGAGACCTTGCTTAGTAATCATCAGTTGC 547
390 euAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLys 406
|||||
548 TTGGCTTTGTGAGGAAAGGACACTGGATAACGGTGGCGAGTAAGAAA 597
407 GluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSe 423
|||||
598 GAAGCCATGATGGTCTGGCTCAGTC.TATAAGAAATACCTGCTTCATGG 646
423 rAlaAlaGlyLysAspAla...AlaLysGlnIleAlaTrpIleLysAspL 439
:::|||||
647 GGAAGCAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696
439 ysLeuLeuHisIleTyrTyrGln.AsnSerIleAspArgLeuLeuVa 455
|||||
697 AGTGGTGGATATCTACTATCAAGACAGCTTTGTTGACAAATATGTGTG 746
455 lGluArgIlePheAlaGlnTyrMetValPro 465
|||||
747 TGAGGAAAAATTTGGTCAGATTGTTGGTCCCA 777

```

seq_name: gb_est28:AL045878

```

seq_documentation_block:
LOCUS AL045878 738 bp mRNA EST 29-FEB-2000
DEFINITION DKFZp434J056_r1 434 (synonym: htes3) Homo sapiens cDNA clone
DKFZp434J056 5', mRNA sequence.
ACCESSION AL045878
VERSION AL045878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 738)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Koehrer, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Koehrer K
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert

```


943 CysAlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaAr 959
 |||||
 927 TTTGCTTTTGTGTCACCAAGACCTGTGAAGAAGAGAGACACAGCTCG 976
 959 glnCysLeuValLysAsnIleAsnValArgGluTyrLeuLysGlnH 976
 |||||
 977 GCAGTGTGTTGTTAAAGAACATCAGCATCCGACGGAGTACATCAACACA 1026
 976 isAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValVal 992
 |||||
 1027 ACCCATGCCACTGAGAATATTGTCTACTGCTGCTGAATATGTGGTT 1076
 993 ProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValG1 1009
 |||||
 1077 CCATACATGATTCACCTCCTAGCCCATGATCTGATTTCCACACGGTCACA 1126
 1009 nAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuG 1026
 |||||
 1127 AGATGTTGATCAACTCGGTGATATAAAGAGTGCCTGTGTTTATGCTTG 1176
 1026 luIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLys 1042
 |||||
 1177 AGTCTTAATGACAAAGATGAACAACACAGCCATCCATTCATGAAGAAG 1226
 1043 MetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAl 1059
 |||||
 1227 ATGGCAGAGAATATCAAGCTAACACAGAGACGCCAGCTCCCGATGAATC 1276
 1059 aLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1076
 |||||
 1277 CAAGACAAATGAAGAACTTTATACGGTTTGTACGGTGGCTGTGTGTGTTA 1326
 1076 leMetSerLysSerThrTyrSerLeuLysSerProLysAspProVal 1092
 |||||
 1327 TAATAGTAAAGTCTGTTGTCATGCAGACTCACCAAGAGCCAGTC 1376
 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThr 1108
 |||||
 1377 CTCCCAATGAAGTTCTTTACACAGCTGAAAGGTAATTTCCCTCACA 1424

seq_name: gb_est101:BG473117

seq_documentation_block: 701 bp mRNA 21-MAR-2001 EST
 LOCUS BG473117
 DEFINITION 602515089F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:464687 5',
 mRNA sequence.
 ACCESSION BG473117
 VERSION BG473117.1 GI:13405392
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 701)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCMI422 row: a column: 24
 High quality sequence stop: 683.
 Location/Qualifiers
 1. 701
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:464687"
 /clone_lib="NIH_MGC_16"

FEATURES

source

/tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GCCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 233 a 128 c 166 g 174 t
 ORIGIN

alignment_scores:

Quality: 919.00 Length: 234
 Ratio: 4.177 Gaps: 2
 Percent Similarity: 94.017 Percent Identity: 75.641

alignment_block:

US-09-512-581-2 x BG473117 ..

Align seg 1/1 to: BG473117 from: 1 to: 701

351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
 |||||
 2 CTCACAGAATATTAAAGGTTAGATCAGCTGATCCAGAGAAGCTATTTCG 51
 367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
 |||||
 52 TCATGATGTCATTGTTACTATATAACAGCTGCCAAGAGGGACCTGGCCT 101
 384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
 |||||
 102 TAGTAATGATCAGCTGCTGGCTTTGTAAGGAAAGACACCTGGATAAA 151
 401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417
 |||||
 152 CGGTGGCAGTAGTAAGAAAGAGCTATGATGGGTCTGGCTCAGCTTTATAA 201
 417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGln Ile 433
 |||||
 202 GAAATACTGCTTCATGTTGAAGCAGGAAGGAAGCTGCAGACGAAAGTC 251
 434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
 |||||
 252 AGCTGGATAAAGACAACTTCTGCATATTTATATCAGACAGCATTTGA 301
 450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 457
 |||||
 302 CGACAAACTGTTGGTAGAGAAAATCTTGTCTCAGTATCTTGTCGCCCA 351
 467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
 |||||
 352 ACCTGGAAACAGAGAGAGAAATGCTTATATTACTTATATGCTAGT 401
 484 LeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAs 500
 |||||
 402 TTGGATCCAAATGCTGTAAGAGCTCTCAACGAAATGTGGAAGTGCAGAA 451
 500 nLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProL 517
 |||||
 452 CATGCTTCGGAGCCATGTACGGAACTATTGGATTTCACACAGCAGCCTA 501
 517 yThrAspAlaSerValLysAlaIlePhe SerLysValMetValIlePh 533
 |||||
 502 CATCAGAGGCTACTGTTCTGCCATGTTCCGAAAAAATGATGACCATAGC 551
 533 rArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 550
 |||||
 552 AAAGAATCTGCCTGTACCCCGGGAAGACACAAGATTACGTGAGAATAATTA 601
 550 hrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566


```

964 sAsnIleAspValArgGluTyrLeuLysGlnHisAlaAlaValSerG 981
|||||
206 AATATAAATGTAAGCGGGAGTATCTGAAGCAGCATGCGACGTGTAGTG 255
|||||
981 lLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
|||||
256 AAAAATTATTGTCTCTCTACAGAGTATGTTGTCATATACAATTCAC 305
|||||
998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLe 1014
|||||
306 CTTTGGCAGCATGACCCAGATATGTCAAAAGTACAGGATATTGAACAAC 355
|||||
1014 uLysAspValLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaL 1031
|||||
356 TAAAGATGTTAAAGAAATGCTTTGGTTGTCTCGAAATATTATAGGCTA 405
|||||
1031 yAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1047
|||||
406 AAAATGAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAATATT 455
|||||
1048 LysGlnThrLysAspAlaGlnGlyProAspAlaLysMetAsnGluLys 1064
|||||
456 AAACAACAAGATGCCCAATGACCATGATGATGACAGAAATGAATGAAA 505
|||||
1064 sLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1081
|||||
506 ACYGTACATGTGTGTGATGTCGATGATGATGATGATGATGATGATGAT 555
|||||
1081 hrThrTyr.SerLeuGluSer 1087
|||||
556 CTACATACNAGTTTGAATCT 576
|||||

```

seq_name: gb_est101:BG532020

```

seq_documentation_block:
LOCUS BG532020 735 bp mRNA EST 03-APR-2001
DEFINITION 602561049F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4698900 5',
mRNA sequence.
ACCESSION BG532020
VERSION BG532020.1 GI:13523558
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1530 row: a column: 13
High quality sequence stop: 727.
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4698900"
/clone_lib="NIH_MGC_61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcccctggcc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'

```

FEATURES
source

adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

BASE COUNT 223 a 141 c 135 g 236 t
ORIGIN

alignment_scores:

Quality: 944.00 Length: 246
Ratio: 4.196 Gaps: 3
Percent Similarity: 91.463 Percent Identity: 78.049

alignment_block:

US-09-512-581-2 x BG532020 ..

Align seg 1/1 to: BG532020 from: 1 to: 735

```

34 ArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerG1 50
|||||
5 AGACTACTGATGTAGTGAACCTTTATGGATATGGATCAGGACTCAGA 54
|||||
50 uGluGluLys.GluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
|||||
55 AGATGAAAGACAGCAGTATCTCCACATAGCCTTGCATCTTGCATCTGAA 104
|||||
67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCy 83
|||||
105 TTCTTCTCAGGAACCCCAATAAGATGTGCGTCTCCTTGTAGCATGTG 154
|||||
83 sLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerP 100
|||||
155 TTTGGCTGATATCTTTCGTATCTATGCCCCAGAGCTCCATATACTTCC 204
|||||
100 roAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116
|||||
205 ATGATAAACTTAAAGGACATATTTTGTGTTATTACCAGACAATTAAGGT 254
|||||
117 LeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuG1 133
|||||
255 TTGAGAGATACAAAGAGTCCACAGTTTAATAGATACCTTTTATTATTAGA 304
|||||
133 uAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAsps 150
|||||
305 GAATTTAGCTTGGTTAAATCATATAACATCTGCTTTGAATTGGAAGATT 354
|||||
150 eAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn 166
|||||
355 GCAATGACATTTTATTACGCTTTTATAGAACTCTCTTCTCAGTGATCAAC 404
|||||
167 AspGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSe 183
|||||
405 AATAGCCACANTAGAGAGGTACAAATGCATGCTAGATTGTGAGTTTC 454
|||||
183 rIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 200
|||||
455 TATCATCATGGAAGGTGATGGAGTTACTTACTCAAGAAATATTGGACTCCATTC 504
|||||
200 euValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAsp 216
|||||
505 TTATTAAACCTCATCTCTGTCACATAAGAACTTAATAAACAGTCTTTTGGC 554
|||||
217 LeuAlaLysAlaLeuLeu.LysArgThrAlaGlnAlaLleGluProTyrI 233
|||||
555 CTTGCAAAAGTGCTATTGACACAGACAGTCCACAGACTATTGAGGCATGCA 604
|||||
233 lerThrThrPheAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249
|||||
605 TTGCCAATTTTTTCAATCAAGTCTCTGTGCTGGGAAGATCATCAGTACGT 654
|||||

```


seq_documentation_block:
 LOCUS AW770571 578 bp mRNA EST
 DEFINITION hl86f11.x1 NCI CGAP-Lu24 Homo sapiens cDNA clone IMAGE:3008877 3'

REFERENCE 1 (bases 1 to 683)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M8780 row: i column: 03
 High quality sequence stop: 607.
 FEATURES
 source
 1. .683
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3598922"
 /clone_lib="NCI_CGAP_Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies, Investigator providing samples: Jeffrey Green, M.D., NIH"
 BASE COUNT 235 a 146 c 138 g 163 t 1 others
 ORIGIN
 alignment_scores:
 Quality: 974.50 Length: 226
 Ratio: 4.575 Gaps: 4
 Percent Similarity: 94.248 Percent Identity: 89.823
 alignment_block:
 US-09-512-581-2 x BE532986 ..
 Align seg 1/1 to: BE532986 from: 1 to: 583
 978 AlaValSerGluLysLeu. LeuSerLeuLeuProGluTyrValValProt 994
 |||||
 3 GCTGTTAGTGAATAATATGTCTCTCTACCAGAGTATGTTGTCAT 52
 994 yThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAsp 1010
 |||||
 53 ATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAAGTACAGGAT 102
 1011 IleGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluIrl 1027
 |||||
 103 ATTGAACAACTTAAGATGTGAAGAATGCCCTTTGTTGTTCTCGAGAT 152
 1027 eLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLysMetV 1044
 |||||
 153 ATTGATGGCTAAATAATGAAACAAACAGCCATCATTTATCAAGAAATGG 202
 1044 aGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
 |||||
 203 TAGAAATATTAAACAGACAAAGAATGCTCAAGGACCATGATGATACAAA 252
 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleLeu 1077
 |||||
 253 ATGAATGAAAAATTGTACACCTGTGTGATGTTCATGCAATCATCATCAT 302
 1077 tSerLysSerThrThrTyrSerLeuSerProLysAspProValLeuP 1094
 |||||
 303 GTCCAAAGACACCATGACAGCTGGAGTCTCTTAAGGACCCCGCTGTC 352
 1094 roAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsn 1110

|||||
 353 CAGCTCGGTTTTTCCACCAGCCTGACAAAGATTTTGTAGTAACACCAAAAT 402
 1111 TyrLeuProGluMetLysSerPhePheThrProGlyLysProLysTh 1127
 |||||
 403 TACCTGCCTCCAGAAATGAATCATATTTTCTACTCTCTGGAACACCTTAAAC 452
 1127 rThrAsnValLeuGluAlaValAsnLysProLeuSerSerAlaGlyLysG 1144
 |||||
 453 AGCCAATGTTCTCGAGCAGTAAATAGCCACTTTTCATCAGCAGCAAAAC 502
 1144 InSerGlnThrLysSerArgMetGluThrValSerAsnAla_SerSe 1160
 |||||
 503 AGTCTCAGACCAAAATCATCAAGATGGAACACTGTGAGCAACCAANGCAG 552
 1160 rSerSerAsn_ProSerSerProGlyArgIleLysGlyArgLeuAspSer 1176
 |||||
 553 CAGCTCCAAACCCCAAGCTCTCTGGAGGATCAGGGAGGCTTGATAGC 602
 1177 SerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLe 1193
 |||||
 603 TCTGAATGGT.CACAGTGAACATGAAGATATACATGTC...TTCCCTTT 648
 1193 uProGlyLysLysSerAspLys 1200
 |||||
 649 CCGCGAAAAAATGACAGAGAGA 670
 seq_name: gb_est88:BF471938
 seq_documentation_block:
 LOCUS BF471938 608 bp mRNA EST 04-DEC-2000
 DEFINITION UI-M-BH3-awt-g-03-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
 UI-M-BH3-awt-g-03-0-UI 5', mRNA sequence.
 ACCESSION BF471938.1 GI:11541121
 VERSION BF471938
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mEST@mail.nih.gov
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
 should be noted that Bento Soares is generating a small number of
 additional specialized non-redundant arrays of BMAP cDNAs whose
 availability will be considered under appropriate and limited
 collaborative arrangements
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1. .608
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH3-awt-g-03-0-UI"
 /clone_lib="NIH_BMAP_M_S4"
 /dev_stage="27-32 days"

FEATURES
source

Location/Qualifiers
 1. .608
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH3-awt-g-03-0-UI"
 /clone_lib="NIH_BMAP_M_S4"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_M_S4 library is a subtracted library of a series,
 ultimately derived from a mixture of individually tagged

403 CTAATAATGAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAT 452
 1047 ilelysglnThrLysaspAlaGlnGlyProAspAlaLysMetAsnG1 1063
 453 ATTAACAACAAGAGATGCCAGGACCATGATGCAAAATGAATGA 502
 1063 uLysLeuThrValCysAspValAlaMetAsnIleIleMetSerLys 1080
 503 AAAAGTACACATGTGTGATGTCATGTCATATCATCATCATCAAGA 552
 1080 erThrThrTyrSerLeuGluSerProLysAspProValLeuPro 1094
 553 GTACTACATACAGTTTGAATCTCTTAAGACCCGGTACTACCA 596

seq_name: gb_est100:BG387646

seq_documentation_block:
 LOCUS BG387646 802 bp mRNA EST 12-MAR-2001
 DEFINITION 602412496F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4521095 5',
 mRNA sequence.

ACCESSION BG387646
 VERSION BG387646.1 GI:13281092
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 802)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 EMAIL: cgapbs-remail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLAM10419 row: p column: 24
 High quality sequence stop: 637.

FEATURES

Location/Qualifiers
 1..802
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4521095"
 /clone_lib="NIH_MGC_92"
 /tissue_type="embryonal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; Oligo-dr primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 270 a 149 c 167 g 216 t
 ORIGIN

alignment_scores:

Quality: 1023.00 Length: 256
 Ratio: 4.316 Gaps: 7
 Percent Similarity: 92.578 Percent Identity: 87.500

alignment_block:

US-09-512-581-2 x BG387646

Align seg 1/1 to: BG387646 from: 1 to: 802

365 AlaIleArgHisPvalIleValSerIleValThrAlaAlaLysLysAs 381
 1 GCTATTAGACATGATGTTATGTCTCAATAGTTACAGCTGCTCAAAAGGA 50

381 pIleLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrL 398
 51 TATTCCTCTGTCATCATCATCTACTTAATTTTGTGAGAGAGAACAT 100
 398 euAspLysArgTTPArgValArgLysGluAlaMetMetGlyLeuAlaGln 414
 101 TAGACAAAACGATGGAGAGTACGCAAGAACCCATGATGGGACTTGGCCAA 150
 415 IleTyrLysLysTyrAlaLeuGlnSer..AlaAlaGlyLysAspAlaAlaL 431
 151 ATTTATAAGAAATATGCTTTACAGTCAGACGCTGGAAAAGATGCTGCAA 200
 431 ysglnIleAlaTTPiLeLysAspLysLeuLeuHisIleTyrTyrGlnAsn 447
 201 AACAGATAGCATGGATCAAGACAAATTGTACATATATATATCAAAAT 250
 448 SerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVa 464
 251 AGTATTGATGATCGACTACTTGTGAACGGATCTTGTCTCAATACATGGT 300
 464 lProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuT 481
 301 TCCTCAATTTTAGAACTACAGAACGGATGAATGCTTATATATTACTTGT 350
 481 YrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTTPLys 497
 351 ATCCACACATGGATTAAATGCTGTGNAAGCATGTAATGACATGTGAAA 400
 498 CysGlnAsnLeuLeuArgHisGlnVal.LysAspLeuLeuAspLeuIleL 514
 401 TGTCAAAATCTGCTCCGACATCAAGTACAAGGATTTGCTTGACTTGATTA 450
 514 ysgln..ProLysThrAspAlaSerValLysAlaIlePheSer..LysValM 530
 451 AGCAAAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAAGTGA 500
 530 etValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMet 546
 501 TGGTTATTACAAGAAATTTACCTGATCCTGTAAAGGCTCAGGATTTTCATG 550
 547 LysLysPheThrGlnValLeuGluAsp.....AspGluLysIleArgLy 561
 551 AAGAAATTCACACAGAGTGTTCAGACAGATGACTGAGGACAAATCAAGAAA 600
 561 sGlnLeuGlu..ValLeuValSerProThrCysSerCys..LysGlnAlaG 577
 601 GCAGTTAGAAAAGTACTTGTGTAGTCCACATGCTCTCTGCCAAAGCGAGCT 650
 577 luGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnPro 593
 651 GACAGTGTGTGCGTGACATTAACCTAGAGTTGGGCAACCCCAACACACT 700
 594 ThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPr 610
 701 ACAAGATC..TTCCTGGCAATGCTCAGCGTCTCTCTGGAGACGATAGCTCT 749
 610 ovalHis 612
 750 GTGCCAT 756

seq_name: gb_est75:BE532986

seq_documentation_block:
 LOCUS BE532986 683 bp mRNA EST 09-AUG-2000
 DEFINITION 601235013F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3598922 5',
 mRNA sequence.

ACCESSION BE532986
 VERSION BE532986.1 GI:9761631
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

201 AGAAGTATCTCTGACACAAATCGTCAAAATTCAGGCTATTAAAAATGATGG 250
833 alargtrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 849
|||||
251 TTTCGATGGCTACTTGAATGAAAAATATACACAGTAAATCAGGAAGTCTCT 300
850 ThrLeuArgLeuLeuThrThrLeuLeuHisSerAspGlyAspLeuThrGl 866
|||||
301 ACCTTAAGATTCCTAACACAATATGTCATAGTGTGAGAGACTTGACAGA 350
866 uGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaG 883
|||||
351 ACAGGGGAAATATTAAACACAGATATGTCAGCTCTGAGACTTGCTGTG 400
883 lYserAlaIleValLysLeuAlaGlnGluProCysTyrHisGluLeuIle 899
|||||
401 CGAGTGCTATTGGAAGCTGGCACAAGAACCCCTGTACCATCAATCATC 450
900 ThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTy 916
|||||
451 ACATTAGAACAAATATCAGCTATGTCATTAGCTATCAACGATGAATGCTA 500
916 rGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgL 933
|||||
501 TCAAGTAGACAAAGTGTGTCGCCAGAACTTCACAAAGCCCTTCCCGTT 550
933 euArg. LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLys. A 949
|||||
551 TACGGGCTTCCACTTGTAGTATATGCAATCTGTGCCCTTGGTGCAAGAAG 600
949 spProValLys...GluArgArgAlaHisAlaArgGlnCysLeuValLys 964
|||||
601 ATCCCTGTAAAGGAGAAGAGAGCTATGCTAGGCAATGTTTGGTGAAAC 650
965 AsnIleAsn.ValArgArgGlu.TyrLeuLysGlnHisAlaAlaValSer 980
|||||
651 ATATATAATGTTCAGGGGAGAGCTATCTGAAGCAGATGGGCTGTAGT 700
981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHi 997
|||||
701 GAGAACTTATGGTCTCTTACCACAGAT.TTGGTGCCCTATATGATTCA 749
997 sLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnL 1014
|||||
750 CTTTGGGGCCTTGGCCCC.....ATTATTTAAGCCGGGATTGACACT 793

seq_name: gb_est75:BE504550

seq_documentation_block: 596 bp mRNA EST 04-AUG-2000
LOCUS BE504550
DEFINITION h258a05.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3212144 3'
similar to TR:094237 094237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA
sequence.
ACCESSION BE504550
VERSION BE504550.1 GI:9706958
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

Seq primer: -400P from Gibco
High quality sequence stop: 458.

FEATURES

Location/Qualifiers

source

1..596

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3212144"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 211 a 103 c 117 g 165 t
ORIGIN

alignment_scores:

Quality: 1027.00 Length: 198
Ratio: 5.187 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-512-581-2 x BE504550 ..

Align seg 1/1 to: BE504550 from: 1 to: 596

897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAs 913
|||||
3 GAAATCATCATATTAGAACAAATATCAGCTATGTCATAGCTATCAACGA 52
913 pGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyL 930
|||||
53 TGAATGCTATCAAGTAAGACAAAGTGTGTCGCCAGAACTTCACAAAGGCC 102
930 euSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946
|||||
103 TTTCCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGT 152
947 AlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuVa 963
|||||
153 GCAAAAGATCCTGTAAAGAGAGAGAGAGCTAGCTAGTCAATGTTTGT 202
963 lYsAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValS 980
|||||
203 GAAAAATATAAATGTAAAGCGGGAGTATCTGAAGCAGCATGCAGCTGTTA 252
980 erGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
|||||
253 GGGAAAAATATTGTCTCTTACCAGAGTATGTTGTTCCATATACAAT 302
997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGl 1013
|||||
303 CACCTTTTGGCAGATGACCCAGATATGTCAAAGTACAGGATATTGAACA 352
1013 nLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMeta 1030
|||||
353 ACTTAAAGATGTTAAAGAAGTCTTTGTTGTTCTCTGGAATATTAAATGG 402
1030 laLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046
|||||


```

53  TTCCACCCTCAGATTACTAACACGATAGTCATAGTGGGATTGA 102
865  hrcGluGlnGlyLysIleSerLysProAspMetSerArgLeuAla 881
|||||
103  CAGAACAGGGAATAGTAACACAGATATGTCACGCTGAGACTGCT 152
|||||
882  AlaglySerAlaIleValLysLeuAlaGlnGluProCystYrHisGlu 898
|||||
153  GCTGGGAGTGCTATTGTGAAGCTGGCACAGGACCTGTACCACGAGAT 202
|||||
898  eileThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGlu 915
|||||
203  CATTACACTGGACAGTACCAGCTGTGTGCATAGCCATCAATGATGAT 252
|||||
915  ystYrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSer 931
|||||
253  GTTATCAAGTCAGCGAGGTGTTCGCTCAGAACTTCACAAAGCCCTTCC 302
|||||
932  ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla 948
|||||
303  CGCTTACGGCTTCCCTTGTGATCATGCGCATCTGTCTTTGTGCCAA 352
|||||
948  sAspProValLysGluArgAlaHisAlaArgGlnCysLeuValLys 965
|||||
353  AGACCCCTGTGAAAGAGAGCGGAGCCCATGCTAGACAGTGTCTGGTGA 402
|||||
965  snIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu 981
|||||
403  ACATCACTGTGAGGAGGAGTACTCTGAAGCAGCATGCAGCTGTGTAGT 452
|||||
982  LysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 998
|||||
453  AAATTATTGCTCTTCTACAGAGATGTGGTTCATATACATTCACCT 502
|||||
998  uLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeu 1015
|||||
503  TTGGGCATGATGCCAGATATGATCAACGTACAGGATATGTAACAACT 552
|||||
1015  ysAspValLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLys 1031
|||||
553  AAGATGTGACAAAGTCCCTGGTTT.GTCTGGAGATATTGATGGGTAA 601
|||||
1032  AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1048
|||||
602  AATGAAACACACAG.CATGCTTTATCAGAAATGTTAGATAATATTA 650
|||||
1048  sGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1065
|||||
651  CAGACCAAGATTTGCTCCAGGACCGATGAT.....CACCACTGAC 691
|||||
1065  euTyrThr.....ValCysAspValAlaMetAsnIleMetSerLys 1079
|||||
692  TGCACAACTGCTCCGCTGTGTGTGTGTGCTGATGCTGATGCTGATG 741
|||||
1080  SerThrTyrSerLeuGluSerPro..... 1088
|||||
742  ...ACACGTACAGCTGGAGTCTAGACCCGGTCCGCTGCTTTACAA 788
|||||
1089  .....LysAspProValLeuProAlaArgPhe.....T 1099
|||||
789  GCGGAGAAATAGTAGTACCAATTCGCTCCAATGATATTTACTCTGG 838
|||||
1099  hrcGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu..... 1112
|||||
839  GCAGTCTCGACAAAACTTT.....CGGACCTTCACCTAATGATGC 879
|||||
1113  .....ProGluMetLysSerPheThrProGlyLysProLysTh 1127
|||||
880  ACGCCGCCCGGAGGAGGACACACGCGACACCCCTCGACACCTATCCA 929
|||||
1127  rThrAsn 1129
|||||
930  CACATCT 936

```

seq_name: gb_est80:BE873840

seq_documentation_block: 848 bp mRNA EST 20-OCT-2000
 LOCUS BE873840 601483992F1 NIH_MGC_69 Homo sapiens cDNA IMAGE:3886511 5',
 DEFINITION mRNA sequence.
 ACCESSION BE873840
 VERSION BE873840.1 GI:10322616
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 848)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov

Tissue Procurement: DCTD/DTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLAM9663 row: g column: 24
 High quality sequence stop: 654.
 Location/Qualifiers

FEATURES
 source

1..848
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3886511"
 /clone_lib="NIH_MGC_69"
 /tissue_type="large cell carcinoma, undifferentiated"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."

BASE COUNT 261 a 163 c 191 g 232 t 1 others
 ORIGIN

alignment_scores:
 Quality: 1061.50 Length: 271
 Ratio: 4.298 Gaps: 6
 Percent Similarity: 91.144 Percent Identity: 86.347

alignment_block:
 US-09-512-581-2 x BE873840 ..

Align seg 1/1 to: BE873840 from: 1 to: 848

750 ProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrPr 766
 |||||
 2 CCTCTCATGAAGAGCTTAGATCCAGCAACCTGGAACATCTCATAACACC 51
 |||||
 766 oLeuValThrIleGlyHisIleAlaLeuAlaProAspGlnPheAla 783
 |||||
 52 ATTTGGTTACTAT.GGTCAATATGCTCTCTTGCACCTGATCAATTGCTG 100
 |||||
 783 laProTyrLysSerTrpValAlaThrPheIleValLysAspLeuMet 799
 |||||
 101 CTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATG 150
 |||||
 800 AsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspG 816
 |||||
 151 AATGATCGGCTTCCAGGAAAAAGACAACATAACTTTGGGTTCAGATGA 200
 |||||
 816 uGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetV 833
 |||||

/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCCTAGCGCCGAGCGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 225 a 129 c 144 g 231 t

ORIGIN

alignment_scores: Quality: 1091.00 Length: 242
Ratio: 4.662 Gaps: 7
Percent Similarity: 96.694 Percent Identity: 95.868

alignment_block:

US-09-512-581-2 x BG432960 ..

Align seg 1/1 to: BG432960 from: 1 to: 729

122 SerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVa 138
2 ACCGCCACAGTTCAATAGGTATTTTATTACTTGAGAACAT.GCTTGGGT 50
138 llySerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheT 155
51 CAAGTCATATACATATGCTTTGAGTTAGATAGCAATGAATTTTCA 100
155 hrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGln 171
101 CCCAGCTATACAGACCTTATTTTTCAGTTATAACAATGGCCACAACTCAG 150
172 LysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluG1 188
151 AAAGTCCATATGACATGCTGATGCTTATGAGCTCTATTTATTGTGAAGG 200
188 yAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValP 205
201 TGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTTAGTAAATCTGGTAC 250
205 roAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu 221
251 CTGCTCATAGAATTAACAAGCAAGCATATGATTGGCAAGGCTTTTA 300
222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAs 238
301 CTGAAGAGGACAGCTCAAGCTATTGAGCCATATATACCAATTTTTTAA 350
238 nGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisV 255
351 TCAGGTTCTGATGCTTGGGAAACATCATATCAGCGATTGTTCAGAGCATG 400
255 al.PheAspLeuLeuGluLeuTyrAsnIleAspSerHisLeuLeuLe 271
401 TCTTTTGACTTAATTTGAGGCTCTACAATATTGATGATTCATTTGCTGCT 450
271 uSerValLeuProGlnLeu.GluPheLysLeuLysSerAsnAspAsnGlu 287
451 CTCTGTTTACCCCGCTTGAATTTAAATTAAGAGCAATGATATAGAG 500
288 GluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAs 304
501 GAGCCCTACAGTTGTAAACTACTGCGAAAAATGTTTGGGCGCAAGGA 550
304 pSerGluLeuAlaSerGln.AsnLysProLeuTrpGlnCysTyr.LeuG1 320
551 TTCAGAAATGGCTTCTCAAAAACAAGCCACTTTTGGCAGTGCTACTTTGGG 600
320 yArgPheAsnAspIleHisValProIle.ArgLeuGluCysVal.LysPh 336

|||||
601 CAGGTTTATGATATCCATGTACCAATTCGCGCTGGAATGTGTGACCAT 650
336 eaLaserHisCysLeuMetAsnHisProAspLeuAla.LysAspLeuThr 352
651 TCTAGCCATTGCTCATGAACCATCTGATTAGCAAAAAGATTTAACA 700
353 GluTyrLeuLys 356
701 GAGTATCTTAAA 712

seq_name: gb_est84:BF142489

seq_documentation_block:

LOCUS BF142489 944 bp mRNA EST 24-OCT-2000
DEFINITION 601788573F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016155 5',
mRNA sequence.

ACCESSION BF142489 GI:10981529

VERSION BF142489.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 944)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9263 row: a column: 20

High quality sequence stop: 620.

FEATURES

Location/Qualifiers

1..944

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:4016155"

/clone_lib="NCI_CGAP_Lu30"

/tissue_type="tumor, metastatic to mammary"

/lab_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; transgenic model WNT-1, expression driven by

MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo

dt. Library constructed by Life Technologies.

Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 277 a 221 c 218 g 228 t

ORIGIN

alignment_scores:

Quality: 1066.00 Length: 319

Ratio: 4.069 Gaps: 7

Percent Similarity: 82.132 Percent Identity: 74.295

alignment_block:

US-09-512-581-2 x BF142489 ..

Align seg 1/1 to: BF142489 from: 1 to: 944

832 MetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyTh 848
|||||
3 ATGGTTTCGATGGCTACTTGAATGAAGAATAATACAGTAAGTCAGGAAC 52
848 rSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuT 865
|||||


```

DEFINITION 602370939F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4479153 5',
            mRNA sequence.
ACCESSION  BG256731
VERSION    BG256731.1  GI:12766547
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 734)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10310 row: m column: 10
            High quality sequence stop: 732.
FEATURES   Location/Qualifiers
            1..734
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4479153"
                /clone_lib="NIH_MGC_92"
                /tissue_type="embryonal carcinoma, cell line"
                /note="organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
                Average insert size 2.5 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT 278 a 154 c 161 g 141 t
ORIGIN
alignment_scores:
    Quality: 1143.50      Length: 241
    Ratio: 4.887          Gaps: 1
Percent Similarity: 97.095 Percent Identity: 95.021
alignment_block:
US-09-512-581-2 x BG256731
Align seg 1/1 to: BG256731 from: 1 to: 734
1078 SerLysSerThrThrThrSerLeuGluSerProLysAspProValLeuPr 1094
1 TCAAAGAGTACTACATACAG.TTGGATCTCTCTAAAGACCGGTACTACC 49
1094 oAlaArgPheThrGlnProAspLysAsnPheSerAsnThrLysAsnT 1111
50 AGCTCGTTCTTCTACACCACTGACAGAAATTCAGTAACACCAAAAT 99
1111 yrLeuProProGluMetLysSerPhePheThrProGlyLysProLysThr 1127
100 ATCTGCCTCTGAAATGAAATCATCTTTTCTACTCTCTGAAAAACCTAAACA 149
1128 ThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGI 1144
150 ACCAATGTTCTAGAGCTGTATACAGCCACTTTCATCAGCGGCAAGCA 199
1144 nSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSers 1161
200 ATCTCAGACCAATCATCAGCAATGGAACCTGTAAACCAATGCAGGACGA 249
1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSer 1177

```

```

250 GCTCAATCCAAGCTCTCTCGAAGAATAAAGGGAGGCTTGATAGTTCT 299
1178 GluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuPr 1194
300 GAAATGGATCACAGTGAATGAAGATTACACAAATGCTTCACCTTTGCC 349
1194 oGlyLysLysSerAspLysArgAspSerAspSerLeuValArgSerGluL 1211
350 GGGGAAAAAAGTGCACAGAGAGACGACTCTGATCTTGTAAGGCTGAAT 399
1211 euGluLysProArgGlyArgLysLysThrProValThrGluGlnGluGlu 1227
400 TGGAGAGGCTTAGAGCAGGAGAAAAAAGCCCGCTCACAGAACAGGAGG 449
1228 LysLeuGlyMetAspLeuThrLysLeuValGlnGluGlnLysProLy 1244
450 AAATTAGGTATGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAA 499
1244 sGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1261
500 AGCAGTCAGCGAAGTCGGAAAGAGGCCATACGGCTTCAGAAATCTGATG 549
1261 lucGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsn 1277
550 AACAGCAGTGGCGCTGAGGAAAAAGAGGCTCAAGAAGATATATTAGAACA 599
1278 ...GluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPr 1293
600 CTGAACGATGACAGAAATAGTCGCCCAAAACAAAGGTTACAAAGAGCGG 649
1293 oProLysProLeuGlyGlyThrProLysGluGluProThrMetLysT 1310
650 CACAAAACCTCTTGGTGGAGGTACACCAAAAGAGAGCCCAACCATGAAA 699
1310 hrSerLysLysGlySerLysLys 1317
700-CTTCTATAAAGGAAGCAACAA 722
seq_name: gb_est100:BG432960
seq_documentation_block:
LOCUS      BG432960              729 bp      mRNA           EST
DEFINITION 602495783F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4609533 5',
            mRNA sequence.
ACCESSION  BG432960
VERSION    BG432960.1  GI:13339466
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 729)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1353 row: e column: 22
            High quality sequence stop: 722.
FEATURES   Location/Qualifiers
            1..729
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4609533"
                /clone_lib="NIH_MGC_75"
                /lab_host="DH10B (T1 phage-resistant)"

```


gls22:AZ185970 + 160.00 245.85 0.0001 886 ! AZ185970 SP_1006_AL_F07_T7A Str
gls22:AZ185970 + 160.00 245.67 0.0001 899 ! AQ269542 HS_2022_AL_H12_T7 CIT
gls22:AZ185970 + 160.00 244.78 0.0002 966 ! BG432918 602495738F1 NIH_MGC_75
gls22:AZ185970 + 160.00 244.66 0.0002 975 ! BF631435 HVSMB0015P03f Hordeum
gls22:AZ185970 + 160.00 244.63 0.0002 978 ! BE612455 601451835F1 NIH_MGC_66
gls22:AZ185970 + 160.00 244.41 0.0002 995 ! BF529509 602043272F1 NCI_CGAP_8
gls22:AZ185970 + 160.00 244.07 0.0002 1023 ! BG529840 602358927F1 NIH_MGC_6
gls22:AZ185970 + 160.00 244.33 0.0002 1086 ! BF631286 HVSMB0015103f Hordeum
gls22:AZ185970 + 160.00 242.34 0.0002 1176 ! BG288980 602381239F1 NIH_MGC_9
gls22:AZ185970 + 160.00 242.01 0.0002 1207 ! BF573791 602077718F1 NIH_MGC_6

seq_name: gb_est88:BF509252

seq_documentation_block:
LOCUS BF509252 791 bp mRNA EST 06-DEC-2000
DEFINITION UI-H-B14-aow-c-07-0-UI-s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3086220 3', mRNA sequence.
ACCESSION BF509252
VERSION BF509252.1 GI:11592550
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 791)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Location/Qualifiers
1..791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site1: Not I; Site2: Eco RI; NCI_CGAP_Sub8
is a subtracted library derived from NCI_CGAP_Sub5. The
NCI_CGAP_Sub8 library had 2.5 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub5 was used
as a tracer in a subtractive hybridization with a driver
comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE
clone ids 2732833-2737415, 3068040-3069191; 25% of the
driver population), a pool of clones from NCI_CGAP_Sub4
(IMAGE clone ids 2723592-2729326; 25% of the driver
population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE IDs
2728969-2733190; 25% of the driver population), and
NCI_CGAP_Sub7 (IMAGE IDs 3069192-3072238, 3081864-3084550
; 25% of the driver population). Subtraction was
performed as previously described [Bonaldo, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches to Facilitate Gene Discovery. Genome Research
6, 791-806].
TAG_LIB=NCI_CGAP_Lu5
TAG_TISSUE=lung
TAG_SEQ=CAAC*

BASE COUNT 270 a 151 c 144 g 226 t
ORIGIN

alignment_scores:

seq_name: gb_est98:BG256731
seq_documentation_block: BG256731 734 bp mRNA EST 13-FEB-2001
LOCUS

Quality: 1225.00 Length: 250
Ratio: 4.940 Gaps: 0
Percent Similarity: 99.200 Percent Identity: 99.200
alignment_block:
US-09-512-581-2 x BF509252 ..
Align seg 1/1 to: BF509252 from: 1 to: 791
894 CysTyrHisGluIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAl 910
|||||
13 TGTACCATGAATCATCATATAGAACATATCATGCTATGTCATTAGC 62
910 aileAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuH 927
|||||
63 TATCAACGATGAATGCTATCAAGTAGACAGTGTGTCGCCAGAACTTC 112
927 isLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCys 943
|||||
113 ACAAGGCGCTTTCCCGTTTACGGCTTCCACTTGAGTATATGCAATCTGT 162
944 AlaLeuCysAlaLysAspProValLysGluArgAlaAlaHisAlaArgG 960
|||||
163 GCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAGAGCTCATGCTAGGCA 212
960 nCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 977
|||||
213 ATGTTGGTGAAATATATAATGTAAGCGGGAGTATCTGAAGCAGCATG 262
977 laAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993
|||||
263 CAGCTGTTAGTAAAAATATTGCTCTTCTTACAGAGATGTTGTTCCA 312
994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
|||||
313 TATCAATTCACCTTTTGGCAGATGACCCAGATATATGTCAAAGTACAG 362
1010 pIleGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluI 1027
|||||
363 TATTGAACAACCTAAAGATGTTAAAGAATGCTCTTGGTTGTTCTGAAA 412
1027 leLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMet 1043
|||||
413 TATTATGGCTAAAAATGAAATATACAGTCACGCTTTTATCAGAAAGATG 462
1044 ValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLy 1060
|||||
463 GTAGAAAAATATTAAACAAACAAAGATGCCAAGGACAGATGATGCAAA 512
1060 sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleM 1077
|||||
513 AATGAATGAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCA 562
1077 etSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeu 1093
|||||
563 TGTCAAGAGTACTCATACAGTTTGGAAATCTCTTAAGACCCCGTACTA 612
1094 ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAs 1110
|||||
613 CCAGCTCGTTTCTTCACTCAACTGACAAGAT.TTCAGTAACACCAAAAA 661
1110 nTyrLeuProProGluMetLysSerPhePheThrProGlyLysProLysT 1127
|||||
662 TTATCTGCTCTGAAATGAAATCATTTTTCACCTCTCTGAAACACCTAAA 711
1127 hrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLys 1143
|||||
712 CAAC.AATGTTCTAGAGCTGTTAAACAGCA.CTTTTCATCAGCAGCAAG 759
seq_name: gb_est98:BG256731

gb_est76:BE612723	162.00	246.67	0.0001	1090	BE612723	601452584F1	NIH_MGC_6	gb_est101:BG492841	161.00	245.02	0.0002	1086	BG492841	602536869F1	NIH_MGC
gb_est80:BE868679	162.00	246.61	0.0001	1095	BE868679	601507854F1	NIH_MGC_7	gb_est101:BG501134	161.00	244.52	0.0002	1113	BG501134	602546367F1	NIH_MGC
gb_est102:BG538263	162.00	246.56	0.0001	1100	BG538263	602566730F1	NIH_MGC_7	gb_est101:BG529642	161.00	244.79	0.0002	1124	BG529642	602558075F1	NIH_MGC
gb_est97:BG117849	162.00	246.48	0.0001	1107	BG117849	602314257F1	NIH_MGC_8	gb_est76:BE910153	161.00	243.22	0.0002	1256	BE910153	601503441F1	NIH_MGC
gb_est97:BG115045	162.00	246.33	0.0001	1120	BG115045	602315477F1	NIH_MGC_8	gb_est76:BE6113921	161.00	242.79	0.0002	1300	BE6113921	601504041F1	NIH_MGC
gb_est93:BG860254	162.00	245.29	0.0002	1218	BF860254	963016A10.y2	C_reinh	gb_est86:BF342615	161.00	242.78	0.0002	1301	BF342615	602013750F1	NCI_CG
gb_est19:BF279708	162.00	245.21	0.0002	1226	BF279708	GA_E00533F16	Gossyp	gb_est86:BF342615	161.00	242.77	0.0002	1301	BF342615	602013750F1	NCI_CG
gb_est95:BF981306	162.00	245.03	0.0002	1244	BF981306	602308052F1	NIH_MGC_8	gb_est92:BF792874	161.00	242.19	0.0002	1364	BF792874	602538937F1	NIH_MGC
gb_est96:BG107670	162.00	243.52	0.0002	1405	BG107670	602277836F1	NIH_MGC_8	gb_est86:BF708736	161.00	242.19	0.0002	1364	BF708736	601890033F1	NIH_MGC
gb_est101:BG519486	162.00	242.78	0.0002	1491	BG519486	602277760F1	NCI_CGAP	gb_est80:BF8104114	161.00	241.65	0.0002	1425	BF8104114	601477115F1	NIH_MGC
gb_est95:BG025954	162.00	242.52	0.0002	1523	BG025954	602292113F1	NIH_MGC_8	gb_est89:BF576555	161.00	241.36	0.0003	1459	BF893934	601437711F1	NIH_MGC
gb_est83:BF131624	162.00	242.21	0.0002	1562	BF131624	601820432F1	NIH_MGC_8	gb_est89:BF576555	161.00	240.54	0.0003	1558	BF576555	602133993F1	NIH_MGC
gb_est85:BF219450	162.00	242.14	0.0002	1570	BF219450	601884243F1	NIH_MGC_8	gb_est100:BG390268	161.00	240.50	0.0003	1563	BG390268	602216232F1	NIH_MGC
gb_est97:BG113909	162.00	243.52	0.0002	1574	BG113909	602284367F1	NIH_MGC_8	gb_est89:BF256684	161.00	240.22	0.0003	1599	BF256684	602370873F1	NIH_MGC
gb_est95:BG031363	162.00	242.11	0.0002	1574	BG031363	602290635F1	NIH_MGC_8	gb_est82:BF038463	161.00	239.97	0.0003	1632	BF388463	601460525F1	NIH_MGC
gb_est95:BG068150	162.00	241.59	0.0002	1642	BF68150	963100G03.y1	C_reinh	gb_est98:BG288709	161.00	239.83	0.0003	1650	BG288709	602385523F1	NIH_MGC
gb_est84:BF204687	162.00	241.53	0.0002	1650	BF204687	963100G03.y1	C_reinh	gb_est86:BF344171	161.00	239.70	0.0003	1668	BF344171	602017413F1	NCI_CG
gb_est100:BG394047	162.00	241.25	0.0003	1687	BG394047	602456205F1	NIH_MGC_1	gb_est92:BF797714	161.00	239.63	0.0003	1677	BF797714	602017652F1	NIH_MGC
gb_est84:BF180004	162.00	240.83	0.0003	1745	BF180004	601806351F1	NCI_CGAP	gb_est97:BG114716	161.00	239.61	0.0003	1680	BG114716	602315386F1	NIH_MGC
gb_est83:BF122002	162.00	240.65	0.0003	1771	BF122002	601756344F1	NCI_CGAP	gb_est91:BF672624	161.00	239.59	0.0003	1683	BF121828	602350823F1	NIH_MGC
gb_est93:BF124728	162.00	240.65	0.0003	1771	BF124728	601761407F1	NCI_CGAP	gb_est91:BF672624	161.00	239.53	0.0003	1691	BF672624	602086916F1	NIH_MGC
gb_est97:BG112855	162.00	240.32	0.0003	1819	BG112855	602285053F1	NIH_MGC_8	gb_est92:BF789391	161.00	239.24	0.0003	1732	BF789391	602104707F1	NCI_CG
gb_est85:BF302581	162.00	240.17	0.0003	1841	BF302581	602032428F1	NCI_CGAP	gb_est83:BF099591	161.00	239.03	0.0003	1759	BF099591	601751611F1	NCI_CG
gb_est94:BF965119	162.00	240.16	0.0003	1842	BF965119	602028893F1	NIH_MGC_8	gb_est84:BF166246	161.00	238.90	0.0003	1778	BF166246	601777007F1	NCI_CG
gb_est76:BE612548	162.00	246.07	0.0003	1856	BE612548	601452060F1	NIH_MGC_8	gb_est89:BF539718	161.00	238.78	0.0003	1796	BF539718	602049929F1	NCI_CG
gb_est93:BF115862	162.00	239.51	0.0003	1942	BF115862	602316842F1	NIH_MGC_8	gb_est95:BG033593	161.00	238.35	0.0004	1859	BG033593	602301717F1	NIH_MGC
gb_est84:BF144133	162.00	239.27	0.0003	1979	BF144133	602178823F1	NCI_CGAP	gb_est91:BF673971	161.00	238.05	0.0004	1905	BF673971	602137356F1	NIH_MGC
gb_est94:BF167018	162.00	239.12	0.0003	2004	BF167018	601774161F1	NCI_CGAP	gb_est98:BG244298	161.00	237.83	0.0004	1939	BG244298	602356005F1	NCI_CG
gb_est91:BF673999	162.00	237.33	0.0004	2315	BF673999	602137389F1	NIH_MGC_8	gb_est89:BF570927	161.00	237.77	0.0004	1948	BF570927	602067125F1	NIH_MGC
gb_est86:BF179716	161.50	248.52	0.0001	877	AQ73461	HS_5482_B1_A01_Sp6	RPC	gb_gss10:BG5246035	160.50	249.07	9.3e-05	732	AQ326035	admb002481B1	CUG_R
gb_est22:AZ200698	161.50	248.43	0.0001	881	AZ200698	SP_1026_A1_A04_Sp6	SP6	gb_est41:AV761512	160.50	248.93	9.5e-05	740	AV761512	nv761512	MDS
gb_est101:BG499864	161.50	248.37	0.0001	891	BG499864	602358874F1	NIH_MGC_59	gb_est101:BG499864	160.50	248.28	0.0001	780	BG499864	602545761F1	NIH_MGC
gb_est93:BF864215	161.50	247.66	0.0001	941	BF864215	963050A04.x1	C_reinh	gb_est86:BF343046	160.50	246.76	0.0001	882	BF343046	602017163F1	NCI_CG
gb_est98:BG346290	161.50	247.30	0.0001	968	BG346290	602360801F1	NCI_CGAP	gb_gss30:AZ675249	160.50	246.11	0.0001	883	AZ675249	ENTKY42TF	Entamoeba
gb_est79:BF797116	161.50	247.03	0.0001	989	BF797116	601464725F1	NIH_MGC_67	gb_est101:BG529461	160.50	246.74	0.0001	929	BG529461	602557833F1	NIH_MGC
gb_est101:BG501043	161.50	246.89	0.0001	1007	BG501043	602546450F1	NIH_MGC_8	gb_est102:BG542951	160.50	245.30	0.0002	982	BG542951	602569343F1	NIH_MGC
gb_est98:BG252634	161.50	246.68	0.0001	1010	BG252634	602366272F1	NIH_MGC_9	gb_est79:BE782836	160.50	244.46	0.0002	1049	BE782836	601472342F1	NIH_MGC
gb_est98:BG291331	161.50	246.38	0.0001	1045	BG291331	602388495F1	NIH_MGC_9	gb_est96:BG107887	160.50	244.59	0.0002	1050	BG107887	602781155F1	NIH_MGC
gb_gss2:CN50165N	161.50	246.35	0.0001	1045	AL106323	Drosophila_melanogaster		gb_est83:BF119347	160.50	244.13	0.0002	1090	BF119947	601758029F1	NCI_CG
gb_est102:BG545612	161.50	246.23	0.0001	1054	BG545612	602572869F1	NIH_MGC_7	gb_est100:BG435095	160.50	243.57	0.0002	1140	BG35095	602507103F1	NIH_MGC
gb_est89:BF531096	161.50	246.12	0.0001	1084	BF531096	602072257F1	NCI_CGAP	gb_est100:BG391723	160.50	243.55	0.0002	1142	BG391723	602418010F1	NIH_MGC
gb_est101:BG479714	161.50	244.76	0.0002	1188	BG479714	602527185F1	NIH_MGC_2	gb_est93:BF865614	160.50	243.44	0.0002	1152	BG5614	963016A05.y1	C_reinh
gb_est89:BF530242	161.50	243.76	0.0002	1287	BF530242	602071417F1	NCI_CGAP	gb_est100:BG429352	160.50	243.26	0.0002	1169	BG436367	602509021F1	NIH_MGC
gb_est86:BF31856	161.50	243.25	0.0002	1341	BF31856	601816123F1	NIH_MGC_5	gb_gss16:AOQ780342	160.50	243.24	0.0002	1171	AQ780342	HS_3107_B1_G12	HR
gb_est86:BF315178	161.50	242.48	0.0002	1427	BF315178	601809594F1	NIH_MGC_1	gb_gss16:AOQ780342	160.50	243.00	0.0002	1194	AQ780342	HS_3184_A2_D07	7C
gb_est92:BF032144	161.50	242.47	0.0002	1428	BF032144	601599575F1	NIH_MGC_1	gb_est100:BG435095	160.50	242.93	0.0002	1200	BG35095	602507102F1	NIH_MGC
gb_est86:BF309252	161.50	242.06	0.0002	1476	BF309252	601887702F1	NIH_MGC_1	gb_est95:BG034865	160.50	242.89	0.0002	1204	BG34865	602353232F1	NIH_MGC
gb_est97:BG115872	161.50	241.99	0.0002	1485	BG115872	602368623F1	NIH_MGC_8	gb_est79:BE784823	160.50	242.56	0.0002	1237	BE784823	601474059F1	NIH_MGC
gb_est93:BF863301	161.50	241.97	0.0002	1487	BF863301	6021487528F1	NIH_MGC_8	gb_est91:BE965081	160.50	241.79	0.0002	1316	BE965081	601638868F1	NIH_MGC
gb_est83:BF813019	161.50	241.89	0.0002	1503	BF813019	963042F06.y1	C_reinh	gb_est97:BF8180352	160.50	241.50	0.0002	1347	BF8180352	602331221F1	NIH_MGC
gb_est89:BG257135	161.50	241.29	0.0003	1571	BF810099	601754666F1	NCI_CGAP	gb_est89:BF574012	160.50	241.14	0.0003	1386	BF574012	602131843F1	NCI_CG
gb_est85:BF214528	161.50	241.05	0.0003	1601	BG257135	602370720F1	NIH_MGC_9	gb_est98:BG249529	160.50	240.97	0.0003	1406	BG249529	602319586F1	NIH_MGC
gb_est100:BG4387830	161.50	240.37	0.0003	1692	BF387830	602412732F1	NIH_MGC_9	gb_est97:BG166139	160.50	240.84	0.0003	1421	BG249529	602340748F1	NIH_MGC
gb_est92:BF782966	161.50	239.67	0.0003	1790	BF782966	602350807F1	NIH_MGC_9	gb_est100:BG421340	160.50	240.40	0.0003	1424	BG421340	602541646F1	NIH_MGC
gb_est72:BE839220	161.50	239.46	0.0003	1794	BF82966	602309050F1	NCI_CGAP	gb_est83:BF118725	160.50	240.22	0.0003	1494	BF118725	601755125F1	NCI_CG
gb_est100:BG4399570	161.50	245.82	0.0001	1003	BG4399570	602495454F1	NIH_MGC_7	gb_est79:BE795690	160.50	240.20	0.0003	1496	BE795690	601590696F1	NIH_MGC
gb_est101:BG432423	161.00	246.01	0.0001	1008	BG432423	602495454F1	NIH_MGC_7	gb_est101:BG482337	160.50	239.68	0.0003	1560	BG482337	602526512F1	NIH_MGC
gb_est89:BF582179	161.00	246.59	0.0001	958	BF582179	602099707F1	NCI_CGAP	gb_est83:BF104133	160.50	239.29	0.0003	1510	BF104133	6021647740F1	NIH_MGC
gb_est92:BF797607	161.00	246.50	0.0001	964	BF797607	602257411F1	NIH_MGC_85	gb_est91:BF683695	160.50	238.46	0.0004	1721	BF687013	602102796F1	NCI_CG
gb_est100:BG4399570	161.00	245.82	0.0001	1008	BG4399570	602495454F1	NIH_MGC_7	gb_est86:BF343046	160.50	238.46	0.0004	1721	BF687013	602102796F1	NCI_CG
gb_est101:BG4399570	161.00	245.82	0.0001	1008	BG4399570	602495454F1	NIH_MGC_7	gb_est89:BF574012	160.50	238.46	0.0004	1721	BF687013	602102796F1	NCI_CG
gb_gss16:AOQ74343	161.00	245.43	0.0001	1046	AQ74343	HS_5387_B2_F03_Sp6	RF	gb_est92:BF791346	160.50	231.08	0.0009	3120	AK014496	Mus_musculus	14
gb_est72:BE265715	161.00	247.48	0.0001	891	BE265715	601193728F1	NIH_MGC_77	gb_hct:AK014496	160.50	231.08	0.0009	3120	AK01449		

9b_est86:BF341305	164.00	251.01	7.3e-05	1010	BF341305	602013244F1	NCI_CGAP	9b_est102:BG537228	163.00	247.71	0.0001	1149	BG537228	602565257F1	NH_MG
9b_est80:BE910534	164.00	250.80	7.5e-05	1027	BE910534	601501073F1	NH_MG	9b_est101:BG537228	163.00	247.48	0.0001	1171	BF951902	602565257F1	NH_MG
9b_est93:BF865122	164.00	250.72	7.6e-05	1034	BF865122	9630507C11.x2	C_rein	9b_est95:BF978747	163.00	247.00	0.0001	1217	BF978747	602149272F2	NH_MG
9b_est100:BG432435	164.00	249.90	8.4e-05	1104	BF878940	HS_3153_A1.G12_T7C	CIT	9b_est82:BF9364532	163.00	246.75	0.0001	1242	BF9364532	601455438F1	NH_MG
9b_est86:BF316672	164.00	247.13	0.0001	1381	BF316672	601903239F1	NH_MG	9b_est84:BF207060	163.00	244.87	0.0002	1445	BF968378	602269456F1	NH_MG
9b_est97:BG164035	164.00	246.98	0.0001	1398	BF164035	602341048F1	NH_MG	9b_est101:BG501044	163.00	244.50	0.0002	1489	BF207060	6021870146F1	NH_MG
9b_est96:BG121816	164.00	246.85	0.0001	1412	BF121816	602282516F1	NH_MG	9b_est96:BG106136	163.00	244.24	0.0002	1520	BG501044	602546451F1	NH_MG
9b_est98:BG250581	164.00	246.45	0.0001	1458	BF250581	602362760F1	NH_MG	9b_est98:BG250581	163.00	243.90	0.0002	1562	BG106136	602289962F1	NH_MG
9b_est78:BF783704	164.00	246.44	0.0001	1460	BF783704	601471101F1	NH_MG	9b_est84:BF164623	163.00	243.82	0.0002	1572	BF783704	602380913F1	NH_MG
9b_est98:BG251709	164.00	245.80	0.0001	1537	BG251709	602363612F1	NH_MG	9b_est84:BF164623	163.00	243.70	0.0002	1588	BF164623	602380913F1	NH_MG
9b_est98:BG288934	164.00	245.75	0.0001	1543	BF288934	602383884F1	NH_MG	9b_est84:BF164623	163.00	243.64	0.0002	1595	BF288934	602380913F1	NH_MG
9b_est98:BF369033	164.00	245.46	0.0001	1580	BF369033	6021843381F1	NH_MG	9b_est84:BF369033	163.00	243.57	0.0002	1604	BF369033	602383884F1	NH_MG
9b_est97:BG162055	164.00	245.15	0.0002	1620	BF162055	601756179F1	NCI_CGAP	9b_est84:BF369033	163.00	243.34	0.0002	1635	BF162055	6021843381F1	NH_MG
9b_est97:BG162055	164.00	245.03	0.0002	1635	BF162055	602345169F1	NH_MG	9b_est84:BF369033	163.00	242.63	0.0002	1731	BF369033	602294834F1	NH_MG
9b_est84:BF137780	164.00	244.63	0.0002	1689	BF137780	601782425F1	NCI_CGAP	9b_est84:BF137780	163.00	242.44	0.0002	1757	BF302409	602031413F1	NCI_CGAP
9b_est97:BG115876	164.00	244.57	0.0002	1697	BF115876	602316868F1	NH_MG	9b_est84:BF137780	163.00	242.32	0.0002	1774	BF122019	602031413F1	NCI_CGAP
9b_est84:BF179980	164.00	244.53	0.0002	1702	BF179980	601806325F1	NCI_CGAP	9b_est84:BF179980	163.00	242.15	0.0002	1799	BF121837	602350836F1	NH_MG
9b_est83:BF101169	164.00	244.47	0.0002	1711	BF101169	601754745F1	NCI_CGAP	9b_est84:BF179980	163.00	241.80	0.0002	1851	BF115864	602316847F1	NH_MG
9b_est84:BF144168	164.00	244.38	0.0002	1723	BF144168	601788866F1	NCI_CGAP	9b_est84:BF179980	163.00	241.74	0.0002	1860	BF292068	602386342F1	NH_MG
9b_est80:BG388280	164.00	244.15	0.0002	1756	BF388280	602413278F1	NH_MG	9b_est84:BF179980	163.00	241.61	0.0002	1870	BF694727	602080806F2	NH_MG
9b_est97:BG121865	164.00	243.81	0.0002	1804	BF121865	602350878F1	NH_MG	9b_est84:BF179980	163.00	241.61	0.0002	1879	BF114698	602080806F2	NH_MG
9b_est84:BF144191	164.00	243.75	0.0002	1813	BF144191	601778895F1	NCI_CGAP	9b_est84:BF179980	163.00	241.61	0.0002	1944	BF345553	602019134F1	NCI_CGAP
9b_est84:BF166248	164.00	243.55	0.0002	1843	BF166248	601777009F1	NCI_CGAP	9b_est84:BF179980	163.00	240.92	0.0003	1986	BF962643	601656009F1	NH_MG
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP
9b_est98:BG261851	164.00	243.41	0.0002	1864	BF261851	602373651F1	NH_MG	9b_est84:BF179980	163.00	240.44	0.0003	2065	BF35923	602051543F1	NCI_CGAP

91	est100:BG401417	167.00	256.92	3.4e-05	945	..	BG401417	602464854F1	N1H_MGC_75	gb_gss30:AZ680982	165.00	255.04	4.3e-05	837	A2680982	ENTL146TR	Entamoeba
91	est3:BF866340	167.00	256.08	3.8e-05	1011	..	BF866440	9630681H2	xl C. reinh	gb_est69:BG035369	165.00	254.36	4.7e-05	884	BE035369	MM06H08	Mm Mesodonta
91	est69:BG039530	167.00	256.05	3.8e-05	1014	..	BG039530	AC06A12	Ac Arabidopsi	gb_est101:BG493577	165.00	254.03	4.9e-05	908	BG493577	60253653F1	N1H_MGC_7
91	est85:BF248321	167.00	252.98	5.7e-05	1309	..	BF248321	601821362F1	N1H_MGC_6	gb_est79:BF871618	165.00	253.66	5.2e-05	935	BF871618	601467475F1	N1H_MGC_7
91	est95:BG026360	167.00	252.88	5.7e-05	1309	..	BG026360	602291730F1	N1H_MGC_6	gb_gss18:AO897460	165.00	253.44	5.3e-05	932	AO897460	HS_3134_A2	G06_TTC
91	est80:BE803481	167.00	252.45	6.1e-05	1355	..	BE893481	601438023F1	N1H_MGC_7	gb_est95:BF976340	165.00	253.18	5.5e-05	972	BF976340	602244471F1	N1H_MGC_7
91	est91:BF693957	167.00	251.41	6.9e-05	1473	..	BF693957	602082536F1	N1H_MGC_8	gb_est97:BG117484	165.00	253.10	5.6e-05	978	BG117484	603347606F1	N1H_MGC_7
91	est95:BF240390	167.00	250.43	7.8e-05	1594	..	BF240390	60105954F1	N1H_MGC_5	gb_est99:BG333926	165.00	252.28	6.2e-05	1045	BG333926	60246064F1	N1H_MGC_7
91	est86:BF345942	167.00	250.25	8.0e-05	1618	..	BF345942	602017946F1	N1H_MGC_5	gb_est80:BE906076	165.00	252.07	6.4e-05	1063	BE906076	601497110F1	N1H_MGC_7
91	est79:BF788182	167.00	249.65	8.7e-05	1698	..	BF788182	601482032F1	N1H_MGC_6	gb_est98:BG259657	165.00	250.04	8.2e-05	1252	BG259657	602380004F1	N1H_MGC_7
91	est98:BG288754	167.00	249.25	9.1e-05	1753	..	BG288754	602385578F1	N1H_MGC_9	gb_est95:BF978734	165.00	249.97	8.3e-05	1259	BF978734	602149244F2	N1H_MGC_7
91	est95:BG032879	167.00	248.80	9.7e-05	1818	..	BG032879	602300462F1	N1H_MGC_8	gb_est80:BF933389	165.00	249.68	8.3e-05	1293	BF933389	601436749F1	N1H_MGC_7
91	est91:BF676053	167.00	248.47	0.0001	1868	..	BF676053	602084025F1	N1H_MGC_8	gb_est84:BF137783	165.00	249.14	8.9e-05	1293	BF137783	601782428F1	N1H_MGC_7
91	est98:BG297242	167.00	248.00	0.0001	1939	..	BG297242	602349451F1	N1H_MGC_9	gb_est95:BF979425	165.00	248.06	0.0001	1447	BF979425	602288255F1	N1H_MGC_7
91	est95:BG027903	167.00	247.96	0.0001	1946	..	BG027903	602294882F1	N1H_MGC_8	gb_est91:BF680542	165.00	248.24	0.0001	1468	BF680542	602155280F1	N1H_MGC_7
91	est82:BF033348	167.00	247.93	0.0001	1951	..	BF033348	601458048F1	N1H_MGC_6	gb_est92:BF792678	165.00	246.79	0.0001	1627	BF792678	602253862F1	N1H_MGC_7
91	est43:AW155198	166.50	257.27	3.3e-05	858	..	AW155198	mg1ae0002B13f	Rice blas	gb_est100:BG393495	165.00	246.77	0.0001	1630	BG393495	602211875F1	N1H_MGC_7
91	est100:BG434803	166.50	255.92	3.9e-05	957	..	BG434803	602507322F1	N1H_MGC_75	gb_est100:BF970000	165.00	246.72	0.0001	1636	BF970000	601588024F1	N1H_MGC_7
91	gss18:AO879037	166.50	254.17	4.9e-05	1102	..	AO879037	HS_3154_A2	H06_TTC	gb_est89:BF572284	165.00	246.68	0.0001	1640	BF572284	602077728F1	N1H_MGC_7
91	est84:BF209842	166.50	253.63	5.2e-05	1151	..	BF209842	601873537F1	N1H_MGC_5	gb_est80:BE988507	165.00	246.68	0.0001	1641	BE988507	601512885F1	N1H_MGC_7
91	gss3:CMS02030	166.50	253.10	5.6e-05	1201	..	AL206205	Tetraodon nigroviridis	gb_est98:BG034342	165.00	246.16	0.0001	1711	BG034342	602302238F1	N1H_MGC_7	
91	est98:BG249535	166.50	252.86	5.7e-05	1224	..	BG249535	602319592F1	N1H_MGC_8	gb_est98:BG0536698	165.00	245.96	0.0001	1717	BG0536698	602370891F1	N1H_MGC_7
91	est98:BG292586	166.50	252.77	5.8e-05	1233	..	BG292586	602389213F1	N1H_MGC_9	gb_est89:BF532390	165.00	245.89	0.0001	1749	BF532390	602074429F1	N1H_MGC_7
91	est82:BF036764	166.50	252.61	5.9e-05	1249	..	BF036764	601458023F1	N1H_MGC_6	gb_est101:BG333926	165.00	245.47	0.0001	1781	BG333926	602528465F1	N1H_MGC_7
91	est95:BF029615	166.50	251.91	6.3e-05	1322	..	BG029615	602296884F1	N1H_MGC_8	gb_est101:BG333926	165.00	245.47	0.0001	1781	BG333926	602528465F1	N1H_MGC_7
91	est101:BG478460	166.50	251.84	6.5e-05	1332	..	BG478460	602523912F1	N1H_MGC_9	gb_est86:BF338144	165.00	245.40	0.0001	1819	BF338144	602037943F1	N1H_MGC_7
91	est98:BG287280	166.50	251.81	6.6e-05	1332	..	BG287280	602523912F1	N1H_MGC_9	gb_est84:BF903632	165.00	244.34	0.0002	1987	BF903632	601808761F1	N1H_MGC_7
91	est85:BF301197	166.50	251.34	7.0e-05	1384	..	BF301197	602029615F1	N1H_MGC_9	gb_gss21:AZ130539	164.50	255.66	3.5e-05	686	AZ130539	OSJNB00104013r	C. coli
91	est83:BF120253	166.50	249.50	8.8e-05	1605	..	BF120253	601756512F1	N1H_MGC_9	gb_est80:BE875356	164.50	255.21	4.3e-05	771	BE875356	601489068F1	N1H_MGC_7
91	est80:BE882186	166.50	249.37	9.0e-05	1622	..	BE882186	601505287F2	N1H_MGC_7	gb_est82:AA849835	164.50	255.05	4.3e-05	771	AA849835	EST192602	Normal
91	est85:BF214482	166.50	249.16	9.2e-05	1650	..	BF214482	601845720F1	N1H_MGC_5	gb_gss3:CMS020M0J	164.50	254.49	4.7e-05	817	AL204580	Tetraodon nigroviridis	N1H_MGC_7
91	est80:BE872484	166.50	248.87	9.6e-05	1659	..	BE872484	601447162F1	N1H_MGC_6	gb_est98:BG382325	164.50	254.28	4.8e-05	831	BG382325	602406919F1	N1H_MGC_7
91	est95:BG027865	166.50	248.77	9.6e-05	1692	..	BG027865	602294882F1	N1H_MGC_8	em_estp101:BF631084	164.50	254.25	4.8e-05	833	BF631084	HVSM62034024F	H. pylori
91	est84:BF180017	166.50	248.77	9.7e-05	1709	..	BF180017	601808364F1	N1H_MGC_8	em_estp101:BG533694	164.50	253.29	5.4e-05	900	BG533694	602562348F1	N1H_MGC_7
91	est94:BF965156	166.50	248.48	0.0001	1743	..	BF965156	602268448F1	N1H_MGC_8	gb_est93:BF965294	164.50	252.85	5.7e-05	932	BF965294	963058607	xl C. coli
91	est100:BG434803	166.50	265.08	1.2e-05	427	..	AF539812	cp57020	xl C. reinh	gb_gss16:AO751459	164.50	252.12	6.2e-05	978	AO751459	HS_5573_B2	H02_Sp6
91	est79:BF942312	166.00	258.55	2.8e-05	723	..	BF942312	601591504F1	N1H_MGC_7	gb_gss16:AO751459	164.50	252.12	6.2e-05	978	AO751459	HS_5573_B2	H02_Sp6
91	est100:BG426696	166.00	257.56	3.1e-05	783	..	BG426696	6023431F1	N1H_MGC_75	gb_est93:BF860077	164.50	251.98	6.4e-05	1000	BF860077	963015504F1	N1H_MGC_7
91	gss3:CMS020CFU	166.00	256.61	3.5e-05	845	..	AL191091	Tetraodon nigroviridis	gb_est96:BG112237	164.50	251.65	6.7e-05	1011	BG112237	602282573F1	N1H_MGC_7	
91	gss35:BO8887	166.00	256.18	3.8e-05	875	..	BO8887	F207-Sp6	IGF Arabidopsis	gb_gss18:AO891386	164.50	251.43	6.9e-05	1045	AO891386	HS_4831_B1	E08_77A
91	est80:BE878848	166.00	254.46	4.7e-05	1005	..	BE878848	601493143F1	N1H_MGC_6	gb_est89:BF537075	164.50	251.20	7.1e-05	1065	BF537075	602048980F1	N1H_MGC_7
91	est100:BG432101	166.00	253.94	5.0e-05	1048	..	BG432101	602409696F1	N1H_MGC_7	gb_est101:BG476360	164.50	251.04	7.3e-05	1079	BG476360	602512861F1	N1H_MGC_7
91	est100:BG429872	166.00	252.68	5.9e-05	1160	..	BG429872	602494850F1	N1H_MGC_7	gb_est102:BG566496	164.50	250.72	7.6e-05	1107	BG566496	602585519F1	N1H_MGC_7
91	est81:BE966970	166.00	252.24	6.2e-05	1202	..	BE966970	601661250F1	N1H_MGC_7	gb_est100:BG425865	164.50	250.70	7.6e-05	1109	BG425865	60249223F1	N1H_MGC_7
91	est98:BG294539	166.00	251.68	6.7e-05	1267	..	BG294539	602391866F1	N1H_MGC_8	gb_est102:BG573646	164.50	249.55	8.8e-05	1216	BG573646	602548668F1	N1H_MGC_7
91	est97:BG114247	166.00	251.59	6.8e-05	1267	..	BG114247	602285812F1	N1H_MGC_8	gb_est92:BF797946	164.50	248.78	9.7e-05	1294	BF797946	602548666F1	N1H_MGC_7
91	est99:BG334906	166.00	250.50	7.8e-05	1383	..	BG334906	602461223F1	N1H_MGC_2	gb_est95:BF978745	164.50	248.74	9.7e-05	1398	BF978745	602149267F2	N1H_MGC_7
91	est31:BF683173	166.00	250.19	8.4e-05	1418	..	BF683173	602139135F1	N1H_MGC_4	gb_est98:BG452799	164.50	248.43	0.0001	1331	BG452799	602365367F1	N1H_MGC_7
91	est85:BF248011	166.00	249.97	8.3e-05	1443	..	BF248011	601859141F1	N1H_MGC_5	gb_est101:BG480437	164.50	248.23	0.0001	1353	BG480437	602529659F1	N1H_MGC_7
91	est89:BF538940	166.00	249.66	8.7e-05	1480	..	BF538940	602051180F1	N1H_MGC_1	gb_gss18:AO879040	164.50	247.13	0.0001	1433	AO879040	HS_3154_A2	H09_77C
91	est86:BF307176	166.00	249.63	8.8e-05	1496	..	BF307176	601891309F1	N1H_MGC_1	gb_est98:BG259896	164.50	247.13	0.0001	1478	BG259896	602393424F1	N1H_MGC_7
91	est95:BG036535	166.00	249.39	9.0e-05	1512	..	BG036535	602326303F1	N1H_MGC_9	gb_est86:BF342456	164.50	246.55	0.0001	1549	BF342456	602013962F1	N1H_MGC_7
91	est77:BE618355	166.00	247.95	0.0001	1698	..	BE618355	601462752F1	N1H_MGC_6	gb_est93:BF959303	164.50	246.53	0.0001	1552	BF959303	963010101	yl C. coli
91	est95:BG024018	166.00	247.84	0.0001	1714	..	BG024018	602303237F1	N1H_MGC_8	gb_est98:BG028429	164.50	246.34	0.0001	1575	BG028429	602294335F1	N1H_MGC_7
91	est83:BF104022	166.00	246.40	0.0001	1924	..	BF104022	601754636F1	N1H_MGC_8	gb_est87:BF381948	164.50	246.27	0.0001	1584	BF381948	601812646F1	N1H_MGC_7
91	est73:BE304226	165.50	257.70	3.1e-05	723	..	BE304226	601085045F1	N1H_MGC_1	gb_est91:BF680508	164.50	246.16	0.0001	1598	BF680508	602155233F1	N1H_MGC_7
91	est100:BG430394	165.50	255.61	4.0e-05	856	..	BG430394	602501207F1	N1H_MGC_75	gb_est93:BF859516	164.50	246.15	0.0001	1600	BF859516	963002G01	y4 C. coli
91	est93:BF866439	165.50	255.46	4.1e-05	866	..	BF866439	603609203	xl C. reinh	gb_est86:BF344858	164.50	245.96	0.0001	1625	BF344858	602014155F1	N1H_MGC_7
91	est80:BE910542	165.50	254.75	4.5e-05	917	..	BE910542	601501084F1	N1H_MGC_70	gb_est84:BF260524	164.50						

90	est100:BG395380	+ 171.00	259.25	2.5e-05	1353	BG395380	602457925F1	NH_MGC_1
91	est98:BG288769	+ 171.00	258.66	3.7e-05	1419	BF857699	602385955F1	NH_MGC_9
92	est93:BF858175	+ 171.00	257.13	3.8e-05	1605	BF858175	963001D02.Y1	C. reinh
93	est85:BF303904	+ 171.00	256.49	3.6e-05	1690	BF303904	601886526F2	NH_CGAP_1
94	est89:BF520356	+ 171.00	255.92	3.9e-05	1770	BF520356	602041846F1	NH_CGAP_1
95	est37:BV554832	+ 170.50	274.76	3.5e-06	362	AV554832	AV554832	Arabisidopsis
96	est77:BB640700	+ 170.50	265.01	1.2e-05	795	BG407000	Cr12.L.D05.SP6	Neurospora
97	est100:BG435340	+ 170.50	262.51	1.7e-05	972	BG435340	602508155F1	NH_MGC_79
98	est101:BG492568	+ 170.50	261.44	1.9e-05	1059	BG492568	602575554F1	NH_MGC_7
99	est101:BG492568	+ 170.50	260.52	2.2e-05	1141	BG492568	602575554F1	NH_MGC_5
100	est102:BG531713	+ 170.50	260.50	2.2e-05	1143	BG531713	602594191F1	NH_MGC_7
101	est86:BF344848	+ 170.50	258.26	2.9e-05	1369	BF344848	602041139F1	NH_CGAP_7
102	est102:BG547851	+ 170.50	258.05	3.0e-05	1392	BG547851	602576148F1	NH_MGC_7
103	est85:BF2113348	+ 170.50	258.02	3.0e-05	1395	BF2113348	601864531F1	NH_MGC_4
104	est80:BE900174	+ 170.50	257.99	3.0e-05	1399	BE900174	601867309F1	NH_MGC_4
105	est86:BF344408	+ 170.50	257.68	3.0e-05	1411	BF344408	602014794F1	NH_CGAP_7
106	est91:BF689763	+ 170.50	257.39	3.1e-05	1433	BF689763	602166799F1	NH_MGC_4
107	est97:BG115861	+ 170.50	255.81	3.9e-05	1667	BG115861	602316839F1	NH_MGC_8
108	est85:BF242974	+ 170.50	255.56	4.1e-05	1702	BF242974	601877746F1	NH_MGC_5
109	est97:BG121841	+ 170.50	255.47	4.1e-05	1714	BG121841	602350084F1	NH_MGC_9
110	est84:BF164694	+ 170.50	255.38	4.2e-05	1726	BF164694	601772491F1	NH_CGAP_7
111	est83:BF131611	+ 170.50	255.06	4.3e-05	1772	BF131611	601820418F1	NH_MGC_5
112	est85:BF2117039	+ 170.50	255.03	4.4e-05	1776	BF2117039	602318825F1	NH_MGC_5
113	est96:BG1107169	+ 170.50	254.86	4.4e-05	1800	BG1107169	602279807F1	NH_MGC_8
114	est89:BF570941	+ 170.50	254.38	4.7e-05	1871	BF570941	602076141F1	NH_MGC_6
115	gs29:AZ603105	+ 170.00	267.79	8.5e-06	593	AZ603105	1M0422010F	Mouse.10kb
116	est101:BG500734	+ 170.00	264.65	1.3e-05	764	BG500734	602547144F1	NH_MGC_60
117	est101:BG506172	+ 170.00	264.01	1.4e-05	804	BG506172	601860114F1	NH_MGC_6
118	est100:BG400281	+ 170.00	262.99	1.6e-05	873	BG400281	602464674F1	NH_MGC_75
119	est97:BG177349	+ 170.00	261.40	1.9e-05	977	BG177349	602313955F1	NH_MGC_8
120	est100:BG428269	+ 170.00	260.60	2.2e-05	1068	BG428269	602499399F1	NH_MGC_7
121	gs16:AB908547	+ 170.00	258.94	2.6e-05	1210	AB908547	601491967F1	NH_MGC_7
122	gs16:AO780315	+ 170.00	258.91	2.6e-05	1213	AO780315	HS.3184.A2.B04.T7C	re
123	est84:BF185542	+ 170.00	258.85	2.7e-05	1219	BF185542	601814642F1	NH_MGC_5
124	est89:BF185540	+ 170.00	256.37	3.7e-05	1489	BF185540	6020184637F1	NH_CGAP_7
125	est89:BF185621	+ 170.00	254.13	4.9e-05	1780	BF185621	602316857F1	NH_MGC_8
126	est97:BG115870	+ 170.00	253.12	5.6e-05	1935	BG115870	602316857F1	NH_MGC_8
127	est85:BF236327	+ 170.00	252.50	6.0e-05	2033	BF236327	602025961F1	NH_CGAP_7
128	gs22:AZ2020335	+ 169.50	262.19	1.7e-05	870	AZ2020335	SP_0101.A1.H01.SP6E	st
129	est101:BG500744	+ 169.50	262.13	1.7e-05	874	BG500744	602547156F1	NH_MGC_60
130	gs22:AZ2196657	+ 169.50	261.66	1.9e-05	908	AZ196657	SP_1032.A2.F09.T7A	st
131	gs831:AZ685229	+ 169.50	261.48	1.9e-05	921	AZ685229	ENTMV78TR	Entamoeba histolytica
132	est76:BB643403	+ 169.50	261.29	1.9e-05	935	BG494303	602539424F1	NH_MGC_58
133	est76:BB616265	+ 169.50	260.23	2.2e-05	998	BE616265	601281689F1	NH_MGC_39
134	est89:BF583260	+ 169.50	260.28	2.2e-05	1019	BF583260	602101723F1	NH_CGAP_7
135	est101:BG493611	+ 169.50	258.26	2.9e-05	1194	BG493611	602546775F1	NH_MGC_6
136	est81:BD963961	+ 169.50	257.56	3.1e-05	1263	BD963961	601657310F1	NH_MGC_6
137	est95:BG026974	+ 169.50	257.18	3.3e-05	1303	BG026974	602594074F1	NH_MGC_6
138	est101:BG529438	+ 169.50	256.49	4.1e-05	1383	BG529438	602557812F1	NH_MGC_6
139	est85:BF214489	+ 169.50	255.44	4.1e-05	1493	BF214489	601845732F1	NH_MGC_5
140	est98:BG291060	+ 169.50	254.97	4.4e-05	1557	BG291060	602386859F1	NH_MGC_9
141	est97:BG121842	+ 169.50	254.67	4.6e-05	1594	BG121842	602350842F1	NH_MGC_9
142	est93:BF867461	+ 169.50	254.24	4.7e-05	1637	BF867461	963091F03.Y1	C. reinh
143	est83:BF120664	+ 169.50	254.21	4.8e-05	1655	BF120664	601758246F1	NH_CGAP_7
144	est85:BF236743	+ 169.50	254.13	4.9e-05	1665	BF236743	602028039F1	NH_CGAP_7
145	est89:BF572271	+ 169.50	254.12	4.9e-05	1667	BF572271	602077708F1	NH_MGC_6
146	est88:BF529968	+ 169.50	253.30	5.4e-05	1781	BF529968	602391005F1	NH_MGC_6
147	est89:BF581599	+ 169.50	253.21	5.5e-05	1793	BF581599	602099359F1	NH_CGAP_7
148	est89:BF570913	+ 169.50	253.09	5.6e-05	1811	BF570913	602076106F1	NH_MGC_6
149	est98:BG2966407	+ 169.50	252.42	5.9e-05	1875	BG2966407	602394287F1	NH_MGC_4
150	est85:BF659049	+ 169.50	252.42	6.1e-05	1911	BF659049	602184407F1	NH_MGC_4
151	est85:BF242967	+ 169.50	250.77	7.5e-05	2184	BF242967	601877734F1	NH_MGC_5
152	est97:BG115883	+ 169.50	250.46	7.8e-05	2238	BG115883	602316866F1	NH_MGC_6
153	gs33:CN5028963	+ 169.00	261.20	2.0e-05	880	AL185963	Tetraodon nigroviridis	
154	est41:AV753588	+ 169.00	260.45	2.2e-05	935	AV753588	ENTV53588	NPD Homo sapiens
155	gs30:AZ681834	+ 169.00	260.20	2.2e-05	954	AZ681834	ENTV5377R	Entamoeba histolytica
156	est76:BB573650	+ 169.00	259.84	2.3e-05	982	BB573650	601333688F2	NH_CGAP_1
157	est101:BG477479	+ 169.00	258.63	2.7e-05	1082	BG477479	602521531F1	NH_MGC_2
158	gs33:CN50207H	+ 169.00	257.39	3.2e-05	1196	AZ206342	Tetraodon nigroviridis	
159	est89:BF581613	+ 169.00	257.26	3.3e-05	1209	BF581613	602099599F1	NH_CGAP_7
160	est95:BG033462	+ 169.00	256.71	3.5e-05	1263	BG033462	602301851F1	NH_MGC_6
161	est97:BG178451	+ 169.00	255.98	3.9e-05	1340	BF178451	602330249F1	NH_MGC_8
162	est92:BF793178	+ 169.00	255.92	3.9e-05	1346	BF793178	602255201F1	NH_MGC_8

91	est15:AT064123	+	334.50	545.14	3.0e-21	675	!	AI064123	GH04388.5prime	GH Dros	gb_est33:BF859150	+	190.00	291.56	4.0e-07	1343	!	BF859150	963001C06.x4	C. re
91	est18:BF070641	-	339.00	536.57	7.0e-22	219	!	BE706461	RC1-H70268-170700-019		gb_est33:BF125422	+	190.00	289.43	5.3e-07	1759	!	BF125422	601763548F1	NIH_MG
91	est68:BF597655	+	323.00	547.19	2.3e-21	270	!	BE597655	R1EN055 RIKEN full-1e		gb_est33:BF165623	+	190.00	288.36	6.1e-07	1739	!	BF165623	602345106F1	NIH_MG
91	est96:BG070903	+	324.00	540.37	5.6e-21	358	!	BG070903	H0101E11-3 NIA Mouse		gb_est33:BF101057	+	189.50	286.11	7.9e-07	1916	!	BF101057	601745621F1	NCI_CG
91	est70:BE090826	+	327.00	540.20	5.6e-21	238	!	BE090826	PM2-BT0723-200300-001		gb_est33:BF180633	+	189.50	286.18	8.0e-07	1936	!	BF180633	601808742F1	NCI_CG
91	est30:AU0117250	+	322.00	532.17	1.6e-20	348	!	AU0117250	AU017250 Mouse two-cell		gb_est100:BG420926	+	189.00	292.92	3.4e-07	1050	!	BG420926	602451137F1	NIH_MG
91	est35:BG0831702	+	321.00	530.52	1.5e-20	295	!	BG0831702	BB031702 RIKEN full-1e		gb_est89:BF570080	+	188.50	284.42	1.0e-06	1945	!	BF570080	602185905F1	NIH_MG
91	est41:AV749063	+	321.00	530.05	2.1e-20	360	!	AV749063	AV749063 NPC Homo sapi		gb_est86:BF345572	+	188.50	284.13	1.0e-06	1992	!	BF345572	602019164F1	NCI_CG
91	est86:BF364559	-	317.00	519.04	8.6e-20	306	!	BF364559	PM2-NN1083-280500-003		gb_est97:BF124754	+	187.00	284.96	9.4e-07	1625	!	BF124754	602345138F1	NIH_MG
91	est96:BG083741	+	315.00	519.30	8.3e-20	377	!	BG083741	H3091E11-5 NIA Mouse		gb_est93:BF165643	+	187.50	285.02	9.3e-07	1510	!	BF165643	602176144F1	NCI_CG
91	est95:BF133147	+	308.00	500.76	8.9e-19	691	!	BF133147	CM3-CN0092-251000-410		gb_est85:BF320341	+	186.50	288.53	9.9e-07	1063	!	BF320340	6020361026F1	NCI_CG
91	est95:BF133147	+	303.00	504.30	5.7e-19	245	!	BF133147	CM3-CN0092-251000-410		gb_est98:BF345311	+	186.00	285.30	9.0e-07	1288	!	BF345311	6020361026F1	NCI_CG
91	est31:AU687846	+	303.00	497.58	1.3e-18	421	!	AU687846	nv09811.s1 NCI_CGAP		gb_est100:BG436135	+	186.00	283.75	1.1e-06	1459	!	BG436135	602508536F1	NIH_MG
91	est10:AU681349	+	297.00	486.65	1.5e-18	200	!	AU681349	nv09811.s1 NCI_CGAP		gb_est80:BF906514	+	186.00	281.98	1.4e-06	1683	!	BF906514	601498379F1	NIH_MG
91	est19:AF176317	+	291.50	479.75	1.3e-17	368	!	AF176317	EST00039 HeLa SRG		gb_est98:BF3283942	+	185.00	279.90	9.2e-07	1227	!	BF3283942	602408112F1	NIH_MG
91	est55:BF032371	+	288.00	478.18	1.6e-17	259	!	BF032371	BB032371 RIKEN full-1e		gb_est84:BF138664	+	185.50	280.58	1.6e-06	1760	!	BF138664	601781845F1	NCI_CG
91	est64:BB423438	+	286.00	473.91	2.8e-17	278	!	BB423438	BB423438 RIKEN full-1e		gb_est85:BF244584	+	185.50	280.18	1.7e-06	1817	!	BF244584	601862601F1	NIH_MG
91	est41:AW060045	-	282.00	464.23	9.7e-17	351	!	AW060045	wz81809.x1 NCI_CGAP		gb_est89:BF569058	+	185.50	279.54	1.9e-06	1914	!	BF569058	602186424F1	NIH_MG
91	est68:BF083957	+	280.00	455.36	6.5e-17	208	!	BF083957	BB589397 RIKEN full-1e		gb_est89:BF179987	+	185.50	279.54	1.9e-06	2047	!	BF179987	602186424F1	NIH_MG
91	est104:FI2354	+	276.00	455.36	3.0e-16	316	!	FI2354	HSC38H11 normalized inh		gb_gss20:AZ020428	+	185.00	292.01	2.9e-07	557	!	AZ020428	RPCI-27-335H4	TV RP
91	est21:AF1532316	+	274.50	445.89	1.0e-15	552	!	AF1532316	SD03779.5prime SD Dros		gb_est102:BG546853	+	185.00	294.92	3.4e-07	606	!	BG546853	602574035F1	NIH_MG
91	est82:BF007848	+	274.00	453.42	5.9e-16	281	!	BF007848	1589716 Amblyomma amer		gb_est97:BF115853	+	185.00	279.18	2.0e-06	1840	!	BF115853	602316828F1	NIH_MG
91	est84:BF149351	+	274.00	450.25	3.8e-16	363	!	BF149351	28.6 Human Epidermal k		gb_est95:BG035008	+	184.50	281.98	1.4e-06	1371	!	BG035008	602325282F1	NIH_MG
91	est54:AW931219	+	274.00	444.09	1.3e-15	596	!	AW931219	RDEIV128 Human rhabdomy		gb_est89:BF525803	+	184.50	280.60	1.6e-06	1532	!	BF525803	602069350F1	NCI_CG
91	est106:BF83765	+	268.00	446.80	9.0e-16	211	!	BF83765	KK2745F Human fetal hear		gb_est89:BF120646	+	184.00	279.92	2.0e-06	1600	!	BF120646	6017584228F1	NCI_CG
91	est83:BF085557	+	262.00	435.24	4.0e-15	236	!	BF085557	MR2-GN0030-190900-008		gb_est97:BF165641	+	184.00	279.92	2.0e-06	1600	!	BF165641	602345136F1	NIH_MG
91	est54:AW996752	+	259.50	420.98	2.5e-14	529	!	AW996752	QV3-BN0047-230200-100		gb_est102:BG575872	+	184.00	278.86	2.0e-06	1647	!	BG575872	602598690F1	NIH_MG
91	est2:AA074000	+	259.00	419.14	3.1e-14	573	!	AA074000	mm82408.71 Strategene		gb_est98:BF293855	+	183.50	280.27	1.7e-06	1373	!	BF293855	602390369F1	NIH_MG
91	est68:BF858516	-	256.00	423.81	1.7e-14	261	!	BF858516	BB585916 RIKEN full-1e		gb_est100:BG426850	+	183.00	284.72	9.7e-07	896	!	BG426850	602493031F1	NIH_MG
91	est84:AA2861628	-	253.00	405.25	1.9e-13	773	!	AA2861628	ZM0168C03F Mouse 10kb		gb_est100:BG404831	+	183.00	278.12	2.3e-06	1525	!	BG404831	6024240154F1	NIH_MG
91	est55:CN505848	+	248.50	396.53	5.7e-13	844	!	CN505848	AL334529 Tetraodon nigroviridis		gb_est100:BF612139	+	183.00	276.83	2.7e-06	1692	!	BF612139	963021A10.x1	C. re
91	est71:BF050448	+	246.00	391.63	1.1e-12	890	!	BF050448	Tetraodon nigroviridis		gb_gss12:AG25282	+	182.50	289.81	5.0e-07	555	!	AG25282	RPCI-11-367B23	TV R
91	est68:BF8586125	+	245.00	404.79	2.0e-13	269	!	BF8586125	BB586125 RIKEN full-1e		gb_est100:BG452897	+	182.50	282.14	1.3e-06	1030	!	BG452897	60249249F1	NIH_MG
91	est31:AA090343	+	243.00	394.29	7.6e-13	477	!	AA090343	RPCI-23-445C4.TJ RPCI		gb_est98:BG296116	+	182.00	291.73	3.9e-07	1938	!	BG296116	602393903F1	NIH_MG
91	est4:CN504488	+	242.00	384.77	2.6e-12	896	!	CN504488	AL290618 Tetraodon nigroviridis		gb_est52:AW853357	+	182.00	291.73	3.9e-07	444	!	AW853357	RC1-CT02502-30099-0	
91	est102:BG598664	+	239.00	382.43	3.5e-12	718	!	BG598664	EST503564 CSTS Soltanum		gb_gss16:AO743307	+	182.00	280.69	1.6e-06	1081	!	AO743307	HS.5387.B2.B04.SP6	
91	est106:N58565	+	235.00	383.77	2.9e-12	373	!	N58565	yv55B09.r1 Soares fetal		gb_est98:BG259344	+	182.00	277.88	2.1e-06	1291	!	BG259344	602378422F1	NIH_MG
91	est33:BF037759	+	235.00	377.71	6.4e-12	608	!	BF037759	ZM0132G24R Mouse 10kb		gb_est98:BF709592	+	182.00	277.88	2.1e-06	1356	!	BF709592	602378422F1	NIH_MG
91	est71:BF050448	+	229.00	374.85	9.2e-12	361	!	BF050448	RC1-H70268-280200-015		gb_est98:BF868376	+	182.00	275.66	3.1e-06	1621	!	BF868376	601444337F1	NIH_MG
91	est4:BF0206496	+	228.00	364.16	3.6e-11	696	!	BF0206496	HS.5567.B2.D03.SP6 RPI		gb_est92:BF782951	+	182.00	274.43	3.6e-06	1791	!	BF782951	602109036F1	NCI_CG
91	est99:BG364469	-	225.00	367.37	2.4e-11	382	!	BG364469	dab95d08.y1 NICHD XGC		gb_est97:BG1218759	+	182.00	273.87	4.1e-06	1950	!	BG121839	602350838F1	NIH_MG
91	est20:AG036969	-	225.00	358.03	8.0e-11	757	!	AG036969	RPCI-23-285A19.TJ RPCI		gb_est84:BF179978	+	182.00	272.41	4.7e-06	2030	!	BF179978	601806323F1	NCI_CG
91	est22:AF1604717	+	223.00	363.46	4.0e-11	372	!	AF1604717	vd71a12.y1 Beddington		gb_est97:BG115847	+	181.50	273.88	3.9e-06	1748	!	BF15847	602316814F1	NIH_MG
91	est7:AA399798	+	223.00	360.77	5.6e-11	462	!	AA399798	vd71a12.y1 Beddington		gb_est97:BG115847	+	181.50	273.88	3.9e-06	1748	!	BF15847	602316814F1	NIH_MG
91	est71:BF050448	+	223.00	358.38	7.6e-11	560	!	BF050448	uv67a05.y1 Soares mous		gb_est97:BG115847	+	181.50	273.88	3.9e-06	1748	!	BF15847	602316814F1	NIH_MG
91	est4:BF0206496	+	221.00	355.51	1.1e-10	535	!	BF0206496	uv67a05.y1 NCI_CGAP		gb_est97:BG115847	+	181.50	273.88	3.9e-06	1748	!	BF15847	602316814F1	NIH_MG
91	est29:AL586843	+	220.00	353.38	1.4e-10	556	!	AL586843	AL586843 Strategene		gb_est100:BG400929	+	181.00	278.97	2.0e-06	1083	!	BG400929	602467844F1	NIH_MG
91	est79:BF780744	+	212.50	326.69	4.4e-09	1715	!	BF780744	601469616F1 NIH_MG		gb_est94:BF964548	+	181.00	277.08	2.3e-06	1170	!	BF964548	602267844F1	NIH_MG
91	est106:N58572	+	212.00	346.25	3.6e-10	331	!	N58572	yv55B09.r1 Soares fetal		gb_est101:BG533805	+	181.00	277.08	2.3e-06	1261	!	BG533805	602562478F1	NIH_MG
91	est21:AF1530941	+	209.50	336.65	1.2e-09	510	!	AF1530941	SD017592.5prime SD Dros		gb_est93:BF860716	+	181.00	274.81	3.4e-06	1514	!	BF860716	963017D07.x1	C. re
91	est21:AF1530941	+	209.50	336.62	1.2e-09	511	!	AF1530941	SD017592.5prime SD Dros		gb_est99:BG342563	+	181.00	274.60	3.5e-06	1541	!	BG342563	602374740F1	NIH_MG
91	est84:BF206496	+	202.50	309.15	4.2e-08	1797	!	BF206496	601869704F1 NIH_MG		gb_est89:BF572295	+	181.00	273.51	4.1e-06	1682	!	BF572295	60207743F1	NIH_MG
91	est11:AA749512	+	202.00	338.01	1.0e-09	164	!	AA749512	HGBB0747 Human Glialbl		gb_est94:BF1158995	+	181.00	272.80	4.5e-06	1781	!	BF158995	601766305F1	NCI_CG
91	est4:CN504589	+	202.00	324.99	5.5e-09	468	!	CN504589	AL275166 Tetraodon nigroviridis		gb_est97:BG114717	+	181.00	271.97	5.0e-06	1904	!	BG114717	602315387F1	NIH_MG
91	est85:BF214484	+	201.00	305.82	6.4e-08	1914	!	BF214484	601845723F1 NIH_MG		gb_est102:BG567880	+	180.50	275.23	3.3e-06	1367	!	BG567880	602586708F1	NIH_MG
91	est100:BG388150	+	200.50	308.69	4.5e-08	1427	!	BG388150	602413130F1 NIH_MG		gb									

OM of: US-09-512-581-2 to: EST: * out_format : pfs

Date: Sep 25, 2001 11:28 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

MODEL=framet.p2n.model -DEV=slp
Q/cgn2-1/USPTO_spool/US09512581/runat_25092001_095746_18380/app_query.fasta_1.1470
DB=EST -QWMT-rastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
MINMATCH=0.000 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
DELEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
LIST=1000 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
ALIGN=50 -MODB=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
MAXLEN=200000000 -USER=US09512581_@CGN1_1.4971 -NCPU=6
ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-512-581-2

Query length: 1391

Database: EST *

Database sequences: 10228115

Database length: 431459454

Search time (sec): 2155.920000

score_list:

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
gb_est88:BF509252	+	1225.00	2053.29	3.0e-105	791	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	1143.50	1916.01	1.3e-97	724	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	1091.00	1827.07	1.2e-92	729	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	1066.00	1781.47	9.1e-90	944	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	1061.50	1775.17	9.1e-90	848	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	1027.00	1721.04	9.6e-87	596	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	1023.00	1710.57	3.7e-86	802	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	974.50	1630.32	1.1e-81	683	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	969.00	1622.43	3.0e-81	608	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	953.00	1593.50	1.2e-79	703	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	949.00	1589.14	2.1e-79	578	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	944.00	1577.68	9.2e-79	735	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	932.00	1555.06	1.7e-77	883	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	927.00	1539.97	1.2e-76	1504	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	919.00	1535.88	2.0e-76	701	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	899.00	1498.83	2.3e-74	902	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	893.00	1491.15	6.1e-74	738	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	890.00	1488.46	8.6e-74	608	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	884.00	1473.63	5.8e-73	885	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	852.00	1422.97	3.8e-70	662	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	850.00	1415.96	9.4e-70	886	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	830.00	1387.73	3.5e-68	560	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	828.00	1386.26	4.3e-68	480	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	827.00	1377.47	1.3e-67	910	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	817.00	1366.26	5.5e-67	535	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	816.00	1365.48	6.0e-67	490	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	806.00	1345.65	1.0e-65	746	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	806.00	1341.31	1.4e-65	888	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	797.00	1330.81	5.2e-65	648	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	786.00	1315.29	3.8e-64	470	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	783.00	1307.60	1.0e-63	580	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	770.00	1285.77	1.7e-62	570	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	767.00	1275.07	6.6e-62	896	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	751.00	1247.29	2.3e-60	944	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	747.00	1242.92	4.1e-60	777	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	746.00	1248.45	2.0e-60	434	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	741.00	1239.89	1.0e-60	437	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	739.00	1233.51	1.4e-59	595	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	731.00	1223.33	5.1e-59	423	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	731.00	1222.70	5.5e-59	445	! BF509252 UI-H-B14-aow-c-07-0-UI
gb_est88:BF509252	+	726.00	1211.68	2.3e-58	546	! BF509252 UI-H-B14-aow-c-07-0-UI

gb_est13:AA913927	1.3e-57	717.00	1197.94	1.3e-57	483	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	717.00	1197.87	1.3e-57	486	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	680.50	1133.78	4.9e-54	580	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	678.50	1126.12	1.3e-53	818	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	677.00	1125.48	1.4e-53	751	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	677.00	1128.98	9.1e-54	529	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	673.00	1122.95	2.0e-53	498	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	670.00	1116.47	4.5e-53	557	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	667.00	1115.78	4.9e-53	391	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	667.00	1106.99	1.5e-52	794	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	658.00	1098.61	4.5e-52	436	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	649.00	1084.72	2.7e-51	408	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	645.00	1070.03	1.7e-50	772	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	639.00	1067.14	2.5e-50	429	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	638.00	1063.55	4.0e-50	500	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	634.00	1049.86	2.3e-49	872	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	629.00	1050.96	2.0e-49	403	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	616.50	1026.99	4.4e-48	504	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	616.00	1028.37	3.7e-48	421	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	577.50	955.02	4.5e-44	806	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	576.00	958.94	2.7e-44	479	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	574.50	958.14	3.0e-44	416	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	567.50	942.30	2.3e-43	573	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	553.50	920.25	3.9e-42	864	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	553.00	913.11	9.6e-42	889	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	549.50	904.57	2.9e-41	1024	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	548.00	906.09	2.4e-41	738	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	545.50	902.28	3.9e-41	713	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	531.00	884.44	3.2e-40	414	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	524.00	870.86	2.2e-39	475	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	514.50	848.87	3.6e-38	763	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	513.50	850.28	3.0e-38	594	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	512.50	843.04	7.7e-38	929	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	507.00	846.82	4.7e-38	323	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	503.00	837.89	1.5e-37	384	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	498.00	829.64	4.3e-37	377	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	496.50	816.43	2.3e-36	890	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	494.00	822.82	1.0e-36	378	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	469.00	779.91	2.5e-34	394	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	458.00	765.90	1.5e-33	271	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	456.00	756.24	5.3e-33	449	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	455.00	759.07	3.7e-33	312	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	454.00	752.25	4.6e-33	315	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	452.00	752.98	8.0e-33	338	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	448.00	741.85	3.3e-32	480	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	442.00	737.51	1.8e-32	300	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	430.50	707.60	2.7e-30	543	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	429.50	704.57	4.0e-30	674	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	428.50	708.25	2.5e-30	468	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	419.50	685.23	4.7e-29	936	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	416.00	692.01	2.0e-29	336	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	416.00	681.69	7.5e-29	772	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	410.50	674.61	1.9e-28	644	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	410.00	681.80	7.4e-29	337	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	407.00	677.16	1.3e-28	325	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	395.50	643.70	9.8e-27	1001	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	393.00	633.69	2.7e-27	318	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	391.00	639.08	1.8e-26	785	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	383.50	628.82	6.2e-26	644	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	380.50	630.02	5.6e-26	388	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	375.00	618.48	2.5e-25	464	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	374.00	623.78	1.3e-25	264	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	374.00	617.87	2.7e-25	425	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	373.50	604.79	1.4e-24	1139	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	372.00	618.23	2.6e-25	314	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	370.00	603.47	1.7e-24	785	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	353.00	585.47	1.7e-23	328	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	341.00	557.10	6.5e-22	626	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	340.00	555.41	8.1e-22	626	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	337.00	545.37	2.9e-21	933	! AA913927 0139f02.sl Soares_N
gb_est13:AA913927	+	334.50	546.04	2.7e-21	628	! AA913927 0139f02.sl Soares_N

THIS PAGE BLANK (USPTO)

VERSION KEYWORDS SOURCE

295113.2 GI:6572188
HTG; ca repeat polymorphism.
human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 110042)
Burton, J.

TITLE

Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:2253036.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

COMMENT

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
CFA-175E3 is from the human BAC library described in U-J. Kim et
al. (1996) Genomics 34, 213-218.
VECTOR: pBAC108L
This sequence is the entire insert of clone CFA-175E3.

FEATURES

source

Location/Qualifiers

1..110042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CFA-175E3"
/clone_lib="CT978SK-A1"

misc_feature

28..515

/note="match: STS: Em:B14235; match: GSS: Em:B14235"

repeat_region

181..318

/note="FLAM_A repeat: matches 1..133 of consensus"

repeat_region

328..437

/note="L2 repeat: matches 2092..2206 of consensus"

repeat_region

584..868

/note="AluSg repeat: matches 1..280 of consensus"

misc_feature

1029..1644

/note="match: GSS: Em:B18037; match: STS: Em:B18037"

repeat_region

1194..1496

/note="AluSx repeat: matches 1..298 of consensus"

repeat_region

2176..2209

/note="17 copies 2 mer tt 88 conserved"

repeat_region

2931..3180

/note="AluJo repeat: matches 60..310 of consensus"

repeat_region

4076..4378

/note="AluJo repeat: matches 2..302 of consensus"

repeat_region

4413..4564

/note="MER91B repeat: matches 28..179 of consensus"

repeat_region

4885..5107

/note="AluSg/x repeat: matches 82..301 of consensus"

repeat_region

5182..5484

/note="AluSx repeat: matches 2..304 of consensus"

repeat_region

5632..5799

/note="AluSg/x repeat: matches 121..300 of consensus"

repeat_region

5875..6032

repeat_region

8190..8485

/note="MIR repeat: matches 21..195 of consensus"

misc_feature

complement(8360..8763)

/note="AluJo repeat: matches 1..299 of consensus"

misc_feature

complement(8421..8889)

/note="match: STS: Em:R08275"

misc_feature

8638..9058

/note="match: GSS: Em:AQ547651"

misc_feature

8669..9019

/note="match: GSS: Em:AQ219894"

misc_feature

8895..9305

/note="match: GSS: Em:AQ550076"

misc_feature

9043..9156

/note="match: GSS: Em:AQ347278"

repeat_region

9292..9586

/note="MIR repeat: matches 106..221 of consensus"

repeat_region

9661..9872

/note="Aluub repeat: matches 1..298 of consensus"

repeat_region

10043..10089

/note="MIR repeat: matches 9..231 of consensus"

repeat_region

10220..10541

/note="L2 repeat: matches 2702..2749 of consensus"

repeat_region

12139..12445

/note="AluSg repeat: matches 1..310 of consensus"

repeat_region

12867..13141

/note="AluSg repeat: matches 1..306 of consensus"

repeat_region

13142..13238

/note="AluSx repeat: matches 1..263 of consensus"

repeat_region

13288..13589

/note="Aluub repeat: matches 51..145 of consensus"

repeat_region

13913..14079

/note="AluSg repeat: matches 1..297 of consensus"

repeat_region

14271..14346

/note="L2 repeat: matches 2428..2602 of consensus"

repeat_region

15560..15660

/note="MIR repeat: matches 59..144 of consensus"

repeat_region

15903..16213

/note="L2 repeat: matches 1365..1458 of consensus"

repeat_region

17782..18082

/note="AluSg repeat: matches 2..311 of consensus"

repeat_region

18355..18736

/note="AluSx repeat: matches 1..302 of consensus"

repeat_region

19322..19608

/note="MLT1B repeat: matches 1..390 of consensus"

repeat_region

20518..20553

/note="AluJo repeat: matches 1..289 of consensus"

repeat_region

20822..20933

/note="18 copies 2 mer ac 100 conserved"

repeat_region

20824..20978

/note="FLAM_A repeat: matches 18..129 of consensus"

repeat_region

21715..21834

/note="BC200 repeat: matches 21..176 of consensus"

repeat_region

21720..21829

/note="6 copies 20 mer 79 conserved"

repeat_region

22568..22730

/note="55 copies 2 mer tg 84 conserved"

repeat_region

22791..23055

/note="MIR repeat: matches 87..256 of consensus"

repeat_region

23216..23514

/note="AluSx repeat: matches 1..298 of consensus"

repeat_region

23517..23819

/note="AluSg repeat: matches 1..299 of consensus"

repeat_region

24150..24208

/note="AluSx repeat: matches 1..303 of consensus"

repeat_region

25160..25475

/note="MIR repeat: matches 77..130 of consensus"

repeat_region

26167..26254

/note="MER46B repeat: matches 2..235 of consensus"

repeat_region

26831..26980

/note="MIR repeat: matches 61..151 of consensus"

repeat_region

27721..27992

/note="L1MC5 repeat: matches 7710..7869 of consensus"

repeat_region

5875..6032

/note="AluJo repeat: matches 5..289 of consensus"

BASE COUNT						31945 a	18533 c	18883 g	30840 t	724 others
ORIGIN										
Query Match						0.6%;	Score 30;	DB 83;	Length 100925;	
Best Local Similarity						100.0%;	Pred. No.	0.00023;		
Matches						30;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY 4428						gaaaaatcctttttttttttttttttttttt 4457				
Db 1637						GAAAAATCTTTTTTTTTTTTTTTTTTTT 1608				
RESULT 47										
AC073980/c						AC073980	101539 bp	DNA	PRI	08-DEC-2000
LOCUS						Homo sapiens clone RP11-312H11,	complete sequence.			
DEFINITION						AC073980				
ACCESSION						AC073980.5	GI:11597142			
VERSION						HTG.				
KEYWORDS						human.				
SOURCE						Homo sapiens				
ORGANISM						Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE						1 (bases 1 to 101539)	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS						Waterston,R.H.				
TITLE						The sequence of Homo sapiens clone				
JOURNAL						Unpublished				
REFERENCE						2 (bases 1 to 101539)				
AUTHORS						Waterston,R.H.				
TITLE						Direct Submission				
JOURNAL						Submitted (08-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE						3 (bases 1 to 101539)				
AUTHORS						Waterston,R.H.				
TITLE						Direct Submission				
JOURNAL						Submitted (07-DEC-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE						4 (bases 1 to 101539)				
AUTHORS						Waterston,R.H.				
TITLE						Direct Submission				
JOURNAL						Submitted (08-DEC-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
COMMENT						On Dec 7, 2000 this sequence version replaced gi:11276284. Center project name: H_NH0312H11.				
FEATURES						Location/Qualifiers				
source						1..101539				
						/organism="Homo sapiens"				
						/db_xref="taxon:9606"				
						/clone="RP11-312H11"				
BASE COUNT						34020 a	18550 c	18384 g	30585 t	
ORIGIN										
Query Match						0.6%;	Score 30;	DB 88;	Length 101539;	
Best Local Similarity						100.0%;	Pred. No.	0.00023;		
Matches						30;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY 4428						gaaaaatcctttttttttttttttttttttt 4457				
Db 25994						GAAAAATCTTTTTTTTTTTTTTTTTTTT 25965				
RESULT 48										
HS175E3						HS175E3	110042 bp	DNA	PRI	12-DEC-1999
LOCUS						Human DNA sequence from clone CTA-175E3 on chromosome 22q12.1				
DEFINITION						Contains an exon of a novel gene, ESTs, STSS, CA repeat polymorphism and GSSs, complete sequence.				
ACCESSION						295113				

FEATURES	source
----------	--------


```

RESULT 40
AC021197
LOCUS      AC021197 172105 bp  DNA           HTG           07-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone RP11-798B19, WORKING DRAFT
SEQUENCE 1, 21 unordered pieces.
ACCESSION AC021197
VERSION    AC021197.5 GI:8570404
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Waterston, R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 172105)
            Waterston, R.H.
            Direct Submission
            Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Jun 17, 2000 this sequence version replaced gi:7235351.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0798B19
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembler: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160434 bases at least Q40
Consensus quality: 164353 bases at least Q30
Consensus quality: 166763 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 170105; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1301: contig of 1301 bp in length
1302 1401: gap of unknown length
1402 4060: contig of 2659 bp in length
4061 4160: gap of unknown length
4161 7024: contig of 2864 bp in length
7025 7124: gap of unknown length
7125 9662: contig of 2538 bp in length
9663 9763: gap of unknown length
9763 13070: contig of 3308 bp in length
13071 13170: gap of unknown length
13171 15928: contig of 2758 bp in length
15929 16029: gap of unknown length
16029 19294: gap of unknown length
19295 19395: gap of unknown length
19395 23283: contig of 3889 bp in length
23284 23384: gap of unknown length
23384 27272: contig of 3889 bp in length
27273 27373: gap of unknown length
27373 32535: contig of 5163 bp in length
32536 40106: contig of 7471 bp in length
32636

```

```

FEATURES
        source
            1..172105
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="19"
                /clone="RP11-798B19"
            1..1301
                /note="assembly_name:Contig11"
            1402..4060
                /note="assembly_name:Contig12"
            4161..7024
                /note="assembly_name:Contig13"
            7125..9662
                /note="assembly_name:Contig14"
            9763..13070
                /note="assembly_name:Contig15"
            13171..15928
                /note="assembly_name:Contig16"
            16029..19294
                /note="assembly_name:Contig17"
            19395..23283
                /note="assembly_name:Contig18"
            23384..27272
                /note="assembly_name:Contig19"
            clone_end:SP6
            vector_side:right
            27373..32535
                /note="assembly_name:Contig20"
            32636..40106
                /note="assembly_name:Contig21"
            40207..45656
                /note="assembly_name:Contig22"
            45757..52232
                /note="assembly_name:Contig23"
            52333..59204
                /note="assembly_name:Contig24"
            59305..68409
                /note="assembly_name:Contig25"
            68510..76838
                /note="assembly_name:Contig26"
            76939..88902
                /note="assembly_name:Contig27"
            89003..100235
                /note="assembly_name:Contig28"
            100336..117760
                /note="assembly_name:Contig29"
            117861..142316
                /note="assembly_name:Contig30"
            142417..172105
                /note="assembly_name:Contig31"
BASE COUNT 43255 a 42852 c 43254 g 40653 t 2091 others
ORIGIN

```


JOURNAL

Submitted (22-NOV-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 156840)

REFERENCE

Waterston,R.H.

AUTHORS

Direct Submission

JOURNAL

Submitted (05-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

5 (bases 1 to 156840)

AUTHORS

Waterston,R.

TITLE

Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

JOURNAL

On May 5, 1999 this sequence version replaced gi:4337251.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.edu

----- Summary Statistics

Center project name: H_NH0215P16

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and

sequencing collaboration between the NHGRI Chromosome 7 Mapping

Project (Eric D. Green, Director), John D. McPherson in the

Department of Genetics (Washington University), and the Washington

University Genome Sequencing Center. For additional information

about the map position of this sequence, see

<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send

mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3 6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-21N8, 200 bp overlap.

Actual start of this clone is at base position 111596 of CTB-21N8;

actual end is at base position 156840 of RP11-215P16.

FEATURES

Source

1..156840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2-q22"
/clone="RP11-215P16"
/clone_lib="RPCI-11"
552..938
/rpt_family="L2"
2050..2156
/rpt_family="Alu"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

2158..2427

/rpt_family="Alu"

3750..3781

/rpt_family="(CA)n"

3880..3991

/rpt_family="L2"

3990..4122

/rpt_family="MER1_type"

4464..4601

/rpt_family="Alu"

4602..4621

/rpt_family="(CAA)n"

4622..4768

/rpt_family="Alu"

9706..9767

/rpt_family="AT_rich"

9796..10014

/rpt_family="L1"

10032..10425

/note="match to EST AA780910 (NID:g2840241) ag98f04.sl"

10536..10573

/rpt_family="AT_rich"

10758..11054

/rpt_family="Alu"

11374..11465

/rpt_family="L2"

12331..13075

/rpt_family="L1"

13131..13252

/rpt_family="Alu"

13405..13501

/note="similar to Mus musculus EST AI325828 (NID:g4060257) mf66g02.yl"

<13409..>134473

/gene="WUGSC:H_NH0215P16.1"

join(<13409..13501,49523..49585,67833..67956,

105929..106049,134380..>134473)

/gene="WUGSC:H_NH0215P16.1"

/note="match to BAA74857.1 (PID:g4240157); H_NH0215P16.1"

/codon_start=1

/product="match to KIAA0834"

/protein_id="AAF19245.1"

/db_xref="GI:6624100"

/translation="SLKGLKHNIVLLHDIHTKETLTLVFEVHTDLCQYMDKHPG
GLHPDNVKEFLFOLLRLCLSVIHHRYILHRDLKPCNLISDTGELKLADFLARAKSV
SHYSNEVTVLWYRPPDVLGSGTEYSTCLDMWVGCCIFVEMIQGVAAFPKMKDIQDL
ERIFL"

13903..13945

/rpt_family="(CA)n"

15936..16133

/rpt_family="MIR"

16450..16506

/rpt_family="(CA)n"

17822..17909

/rpt_family="L1"

17935..18069

/rpt_family="L1"

18076..18376

/rpt_family="Alu"

18385..18490

/rpt_family="L1"

18491..18570

/rpt_family="(TA)n"

18575..18717

/rpt_family="L1"

18911..19121

/rpt_family="MIR"

19444..19954

/rpt_family="L2"

20346..20736

/rpt_family="L1"

21973..22023

/rpt_family="MIR"


```

RESULT 33
AC022305
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-276K9 map 15q21, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC022305
VERSION AC022305.2 GI:8247803
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 94409)
AUTHORS Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,
Madan, A., Nesbitt, R., Shaffer, T. and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 94409)
AUTHORS Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,
James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T.
and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2000) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT On Jun 4, 2000 this sequence version replaced gi:6814954.
-----
Center: Multimegabase Sequencing Center
Center code: UHMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leetowen@systemsbiology.org
-----
Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
-----
* NOTE: This record contains 96 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1 865: contig of 865 bp in length
* 866 965: gap of unknown length
* 966 1810: contig of 845 bp in length
* 1811 1910: gap of unknown length
* 1911 2801: contig of 891 bp in length
* 2802 2901: gap of unknown length
* 2902 3779: contig of 878 bp in length
* 3780 3879: gap of unknown length
* 3880 4737: contig of 858 bp in length
* 4738 4837: gap of unknown length
* 4838 5722: contig of 885 bp in length
* 5723 5822: gap of unknown length
* 5823 6703: contig of 881 bp in length
* 6704 6803: gap of unknown length
* 6804 7684: contig of 881 bp in length
* 7685 7784: gap of unknown length
* 7785 8676: contig of 892 bp in length
* 8677 8776: gap of unknown length
* 8777 9672: contig of 896 bp in length
* 9673 9772: gap of unknown length
* 9773 10655: contig of 883 bp in length
* 10656 10755: gap of unknown length
* 10756 11664: contig of 909 bp in length
* 11665 11764: gap of unknown length
*
11765 12657: contig of 893 bp in length
12658 12757: gap of unknown length
12758 13630: contig of 873 bp in length
13631 13730: gap of unknown length
13731 14598: contig of 868 bp in length
14599 14699: gap of unknown length
14700 15576: contig of 878 bp in length
15577 15677: gap of unknown length
15678 16549: contig of 873 bp in length
16550 16649: gap of unknown length
16650 17571: contig of 922 bp in length
17572 17671: gap of unknown length
17673 18543: contig of 872 bp in length
18544 18643: gap of unknown length
18644 19522: contig of 879 bp in length
19523 19622: gap of unknown length
19623 20520: contig of 898 bp in length
20521 20620: gap of unknown length
20621 21523: contig of 903 bp in length
21524 21623: gap of unknown length
21624 22502: contig of 879 bp in length
22503 22602: gap of unknown length
22603 23485: contig of 883 bp in length
23486 23585: gap of unknown length
23586 24468: contig of 883 bp in length
24469 24568: gap of unknown length
24569 25447: contig of 879 bp in length
25448 25547: gap of unknown length
25549 26431: contig of 884 bp in length
26432 26531: gap of unknown length
26532 27406: contig of 875 bp in length
27407 27507: gap of unknown length
27508 28395: contig of 889 bp in length
28396 28495: gap of unknown length
28496 29380: contig of 885 bp in length
29381 29480: gap of unknown length
29481 30379: contig of 899 bp in length
30380 30479: gap of unknown length
30480 31086: contig of 607 bp in length
31087 31186: gap of unknown length
31187 32077: contig of 891 bp in length
32078 32177: gap of unknown length
32178 33054: contig of 877 bp in length
33055 34034: contig of 880 bp in length
34035 34134: gap of unknown length
34135 35023: contig of 889 bp in length
35024 35123: gap of unknown length
35124 36014: contig of 891 bp in length
36015 36114: gap of unknown length
36115 37015: contig of 901 bp in length
37016 37115: gap of unknown length
37116 37998: contig of 883 bp in length
37999 38098: gap of unknown length
38099 38973: contig of 875 bp in length
38974 39073: gap of unknown length
39074 39943: contig of 870 bp in length
39944 40043: gap of unknown length
40044 41019: contig of 876 bp in length
41020 41916: contig of 897 bp in length
41917 42916: gap of unknown length
42917 43026: contig of 910 bp in length
43027 43899: gap of unknown length
43899 43999: contig of 873 bp in length
43900 44000: gap of unknown length
44001 45002: contig of 903 bp in length
45003 45879: contig of 877 bp in length
45880 45979: gap of unknown length
45980 46859: contig of 880 bp in length
46860 46959: gap of unknown length
46960 47863: contig of 904 bp in length

```


Query Match 0.6%; Score 32; DB 80; Length 283512;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;

* be preserved.
* 1 799: contig of 799 bp in length
* 800 899: gap of 100 bp

	source	1..239086 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10" /clone="RP11-138M8" /clone_lib="RPCI-11"	67358 a 47329 c 47356 g 69901 t 7142 others	
	BASE COUNT			
	ORIGIN			
	Query Match	0.6%; Score 32; DB 69; Length 239086;		
	Best Local Similarity	100.0%; Pred. No. 2.le-05;		
	.Matches 32; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	4429	aaaaatcttttttttttttttttttttggg 4460 		
Dd	25367	AAAAATCTTTTTCCTTTTTTTTTTTGGGT 25398		
RESULT 31				
AL365202				
LOCUS	AL365202	283512 bp DNA HTG 23-JAN-2001		
DEFINITION	Homo sapiens chromosome 9 clone RP11-326A8, *** SEQUENCING IN			
PROGRESS ***	32 unordered pieces.			
ACCESSION	AL365202			
VERSION	AL365202.8 GI:11322083			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 283512)			
TITLE	Plumb,B.			
JOURNAL	Direct Submission			
	Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,			
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone			
	requests: clonerequest@sanger.ac.uk			
COMMENT	On Nov 23, 2000 this sequence version replaced gi:10277983. ----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: ba326a8 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 264525 bases at least Q40 Consensus quality: 271473 bases at least Q30 Consensus quality: 276257 bases at least Q20 Insert size: 280412; sum-of-contigs Insert size: 184090; 20.5% error; agarose-fp Quality coverage: 5.15x in Q20 bases; sum-of-contigs Quality coverage: 8.86x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 32 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 42999: contig of 42999 bp in length * 43000 43099: gap of 100 bp * 43100 49755: contig of 6656 bp in length * 49756 49855: gap of 100 bp * 49856 53234: contig of 3379 bp in length * 53235 53334: gap of 100 bp * 53335 70604: contig of 17270 bp in length * 70605 70704: gap of 100 bp			

* as soon as it is available and the accession number will
 * be preserved.

1 1063: contig of 1063 bp in length
 * 1064: gap of unknown length
 * 1164: contig of 1055 bp in length
 * 2219: gap of unknown length
 * 2319: contig of 1123 bp in length
 * 3441: gap of unknown length
 * 3542: contig of 1122 bp in length
 * 4664: gap of unknown length
 * 4764: contig of 1119 bp in length
 * 5883: gap of unknown length
 * 5983: contig of 1042 bp in length
 * 7025: gap of unknown length
 * 7125: contig of 1451 bp in length
 * 8576: gap of unknown length
 * 8676: contig of 1129 bp in length
 * 9805: gap of unknown length
 * 9905: contig of 1712 bp in length
 * 11617: gap of unknown length
 * 11717: contig of 1007 bp in length
 * 12724: gap of unknown length
 * 12823: contig of 1696 bp in length
 * 14519: gap of unknown length
 * 14520: contig of 1095 bp in length
 * 14620: gap of unknown length
 * 15715: contig of 1179 bp in length
 * 15815: gap of unknown length
 * 16994: contig of 1128 bp in length
 * 17094: gap of unknown length
 * 18221: contig of 1082 bp in length
 * 18322: gap of unknown length
 * 19404: contig of 1021 bp in length
 * 20524: gap of unknown length
 * 20625: contig of 1594 bp in length
 * 22219: gap of unknown length
 * 22319: contig of 1049 bp in length
 * 23368: gap of unknown length
 * 23468: contig of 1076 bp in length
 * 24544: gap of unknown length
 * 24644: contig of 1071 bp in length
 * 25715: gap of unknown length
 * 25815: contig of 1502 bp in length
 * 27317: gap of unknown length
 * 27417: contig of 1104 bp in length
 * 28520: gap of unknown length
 * 28620: contig of 1558 bp in length
 * 28621: gap of unknown length
 * 30179: contig of 1200 bp in length
 * 30279: gap of unknown length
 * 31479: contig of 1058 bp in length
 * 31579: gap of unknown length
 * 32637: contig of 1045 bp in length
 * 32737: gap of unknown length
 * 33781: contig of 1425 bp in length
 * 33881: gap of unknown length
 * 35306: contig of 1119 bp in length
 * 35406: gap of unknown length
 * 36523: contig of 1301 bp in length
 * 36626: gap of unknown length
 * 37927: contig of 1300 bp in length
 * 38027: gap of unknown length
 * 39327: contig of 1751 bp in length
 * 39427: gap of unknown length
 * 41177: contig of 2268 bp in length
 * 41277: gap of unknown length
 * 43545: contig of 1081 bp in length
 * 43645: gap of unknown length
 * 43646: contig of 1267 bp in length
 * 44727: gap of unknown length
 * 44826: contig of 1559 bp in length
 * 46033: gap of unknown length
 * 46193: contig of 1776 bp in length
 * 46194: gap of unknown length
 * 47553: contig of 1776 bp in length
 * 47553: contig of 1776 bp in length

49629: gap of unknown length
 * 49729: contig of 1089 bp in length
 * 50818: gap of unknown length
 * 50918: contig of 1133 bp in length
 * 52051: gap of unknown length
 * 52151: contig of 1851 bp in length
 * 54002: gap of unknown length
 * 54102: contig of 1311 bp in length
 * 55413: gap of unknown length
 * 55513: contig of 1557 bp in length
 * 57070: gap of unknown length
 * 57170: contig of 1296 bp in length
 * 58466: gap of unknown length
 * 58566: contig of 1846 bp in length
 * 60412: gap of unknown length
 * 60512: contig of 1731 bp in length
 * 62243: gap of unknown length
 * 62343: contig of 2121 bp in length
 * 64464: gap of unknown length
 * 64564: contig of 2617 bp in length
 * 67181: gap of unknown length
 * 67281: contig of 2004 bp in length
 * 69285: gap of unknown length
 * 69385: contig of 1925 bp in length
 * 71310: gap of unknown length
 * 71410: contig of 3209 bp in length
 * 74619: gap of unknown length
 * 74719: contig of 2727 bp in length
 * 77446: gap of unknown length
 * 77546: contig of 3305 bp in length
 * 80851: gap of unknown length
 * 80951: contig of 3176 bp in length
 * 84127: gap of unknown length
 * 84227: contig of 3811 bp in length
 * 88038: gap of unknown length
 * 88138: contig of 2680 bp in length
 * 90818: gap of unknown length
 * 90918: contig of 3534 bp in length
 * 94452: gap of unknown length
 * 94552: contig of 3087 bp in length
 * 97639: gap of unknown length
 * 97739: contig of 2745 bp in length
 * 100484: gap of unknown length
 * 100584: contig of 2489 bp in length
 * 103073: gap of unknown length
 * 103173: contig of 4040 bp in length
 * 107213: gap of unknown length
 * 107313: contig of 3536 bp in length
 * 110849: gap of unknown length
 * 110949: contig of 2883 bp in length
 * 113832: gap of unknown length
 * 113932: contig of 3996 bp in length
 * 117928: gap of unknown length
 * 118028: contig of 5035 bp in length
 * 123063: gap of unknown length
 * 123163: contig of 4361 bp in length
 * 127524: gap of unknown length
 * 127624: contig of 5132 bp in length
 * 132756: gap of unknown length
 * 132856: contig of 8828 bp in length
 * 141684: gap of unknown length
 * 141784: contig of 7554 bp in length
 * 149338: gap of unknown length
 * 149438: contig of 10601 bp in length
 * 160039: gap of unknown length
 * 160139: contig of 12150 bp in length
 * 172289: gap of unknown length
 * 172389: contig of 19487 bp in length
 * 191876: gap of unknown length
 * 191976: contig of 18819 bp in length
 * 210795: gap of unknown length
 * 210895: contig of 28192 bp in length

Location/Qualifiers

FEATURES


```

1. /organelle="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13"
/clone="RP11-170D1"
/clone_lib="RPC1-11 Human Male BAC"
1. .1001
/note="assembly_fragment"
1102. .2167
/note="assembly_fragment"
2368. 3456
/note="assembly_fragment"
3557. .4923
/note="assembly_fragment"
5024. .6342
/note="assembly_fragment"
6443. 7840
/note="assembly_fragment"
7941. .9039
/note="assembly_fragment"
9140. 10500
/note="assembly_fragment"
10601. .11995
/note="assembly_fragment"
12086. .13609
/note="assembly_fragment"
13710. .15132
/note="assembly_fragment"
15233. .17325
/note="assembly_fragment"
17426. .19093
/note="assembly_fragment"
19194. .21018
/note="assembly_fragment"
21119. .23488
/note="assembly_fragment"
23589. .26648
/note="assembly_fragment"
26749. .30209

```


numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9>
 RP11-509J21 is from the library RPCI-11.2 constructed at the
 Roswell Park Cancer Institute by the group of Pieter de Jong. For
 further details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-509J21. The true
 left end of clone RP11-661K19 is at 114307 in this sequence. The
 true right end of clone RP11-326A8 is at 32194 in this sequence.

FEATURES

source	Location/Qualifiers
	1..163447
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-509J21"
	/clone_lib="RPCI-11.2"
	1..547
repeat_region	/note="L1M4 repeat: matches 4282..4882 of consensus"
repeat_region	1664..1739
repeat_region	/note="AluJ/FLAM repeat: matches 6..81 of consensus"
repeat_region	1819..1872
repeat_region	/note="Alu repeat: matches 243..296 of consensus"
repeat_region	2942..3214
repeat_region	/note="AluSx repeat: matches 39..312 of consensus"
repeat_region	3581..3702
repeat_region	/note="AluJ/FRAM repeat: matches 185..295 of consensus"
repeat_region	3999..4195
repeat_region	/note="MIR repeat: matches 45..262 of consensus"
repeat_region	4258..4550
repeat_region	/note="AluJ repeat: matches 1..293 of consensus"
repeat_region	4797..5012
repeat_region	/note="AluSg/x repeat: matches 86..300 of consensus"
repeat_region	5026..5558
repeat_region	/note="L2 repeat: matches 1710..2278 of consensus"
repeat_region	5978..6102
repeat_region	/note="L2 repeat: matches 2345..2474 of consensus"
repeat_region	6113..6349
repeat_region	/note="MIR repeat: matches 11..259 of consensus"
repeat_region	6560..6861
repeat_region	/note="AluSg repeat: matches 1..302 of consensus"
repeat_region	8760..8975
repeat_region	/note="AluJb repeat: matches 72..287 of consensus"
repeat_region	9395..9659
repeat_region	/note="AluSx repeat: matches 1..284 of consensus"
repeat_region	10786..11196
repeat_region	/note="MLT1J repeat: matches 1..414 of consensus"
repeat_region	11402..11528
repeat_region	/note="MIR repeat: matches 20..149 of consensus"
repeat_region	11632..11659
repeat_region	/note="14 copies 2 mer ac 100% conserved"
repeat_region	11729..11921
repeat_region	/note="MIR repeat: matches 31..221 of consensus"
repeat_region	11943..12062
repeat_region	/note="L1ME3A repeat: matches 5991..6113 of consensus"
repeat_region	12063..12621
repeat_region	/note="MLT2D repeat: matches 1..553 of consensus"
repeat_region	12622..12672
repeat_region	/note="L1ME3A repeat: matches 6113..6164 of consensus"
repeat_region	12742..13045
repeat_region	/note="AluSg repeat: matches 1..305 of consensus"
repeat_region	13420..13511
repeat_region	/note="L1MC2 repeat: matches 6218..6314 of consensus"
repeat_region	13692..20142
repeat_region	/note="L1PA7 repeat: matches 8..6139 of consensus"
repeat_region	20151..20339
repeat_region	/note="L1MC2 repeat: matches 6044..6234 of consensus"
repeat_region	20353..20841
	/note="Tiger3b repeat: matches 701..1205 of consensus"
	20842..21140
	/note="AluJb repeat: matches 1..288 of consensus"
	21141..21896
	/note="Tiger3b repeat: matches 1..701 of consensus"
	21915..22265
	/note="L1MC2 repeat: matches 5643..6004 of consensus"
	22435..22736
	/note="AluSg repeat: matches 1..303 of consensus"
	23083..23379
	/note="AluSx repeat: matches 3..297 of consensus"
	23984..24170
	/note="MIR repeat: matches 49..244 of consensus"
	25343..25466
	/note="AluSg/x repeat: matches 170..293 of consensus"
	25921..26189
	/note="MER58C repeat: matches 4..89 of consensus"
	27058..27825
	/note="L1MB3 repeat: matches 5446..6176 of consensus"
	27881..27928
	/note="MIR repeat: matches 49..96 of consensus"
	28125..28584
	/note="L2 repeat: matches 1992..2489 of consensus"
	28661..28889
	/note="L2 repeat: matches 1643..1880 of consensus"
	29272..29429
	/note="MER5B repeat: matches 1..175 of consensus"
	29485..29731
	/note="MIR repeat: matches 3..262 of consensus"
	31630..32055
	/note="L2 repeat: matches 2099..2545 of consensus"
	32445..32490
	/note="23 copies 2 mer aa 73% conserved"
	32838..33074
	/note="MER20 repeat: matches 2..218 of consensus"
	33177..33580
	/note="MLT1B repeat: matches 1..390 of consensus"
	34501..34800
	/note="AluJ repeat: matches 1..299 of consensus"
	34910..35029
	/note="L2 repeat: matches 1931..2053 of consensus"
	35051..36193
	/note="MER69B repeat: matches 1..1225 of consensus"
	36596..36913
	/note="L1MD repeat: matches 977..1287 of consensus"
	37179..37473
	/note="AluJb repeat: matches 1..307 of consensus"
	37497..37538
	/note="21 copies 2 mer aa 83% conserved"
	37563..37976
	/note="MSTA repeat: matches 1..426 of consensus"
	37977..38328
	/note="MSTA repeat: matches 1..426 of consensus"
	38341..38696
	/note="L1MEC repeat: matches 1639..1982 of consensus"
	38762..38914
	/note="AluJ/FRAM repeat: matches 150..302 of consensus"
	39048..40267
	/note="L1MEC repeat: matches 2189..3119 of consensus"
	40296..41553
	/note="L2 repeat: matches 433..1770 of consensus"
	41554..41851
	/note="AluJb repeat: matches 1..312 of consensus"
	41852..42145
	/note="L2 repeat: matches 1770..2050 of consensus"
	43119..43335
	/note="MIR repeat: matches 6..262 of consensus"
	43646..44014
	/note="MSTA repeat: matches 1..375 of consensus"
	44015..44395
	/note="THE1C repeat: matches 1..371 of consensus"
	44396..44448
	/note="MSTA repeat: matches 375..426 of consensus"

Query Match 0.8%; Score 41; DB 80; Length 195380;
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 tcatatacatatgcttggagtagtagaagaatgaatgaat 523
 |||||
 DB 170711 TCATATACATATGCTTTGGAGTTAGTAGAATGAATGAAT 170751

RESULT 24
 AC007038 99030 bp DNA PRI 30-SEP-2000
 LOCUS Homo sapiens BAC clone RP11-260M2 from 2, complete sequence.
 DEFINITION AC007038
 ACCESSION AC007038
 VERSION AC007038.3 GI:5708475
 KEYWORDS HTG
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 99030)
 Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 REFERENCE 2 (bases 1 to 99030)
 COURTNEY, L., WOHLDMANN, P. and KETTERMAN, M.
 TITLE The sequence of Homo sapiens BAC clone RP11-260M2
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 99030)
 WATERSTON, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 99030)
 WATERSTON, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (07-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 99030)
 WATERSTON, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Aug 8, 1999 this sequence version replaced gi:4580502.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0260M02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this

sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-436C12, 200 bp overlap; the clone sequenced to the right is RP11-396J78. Actual start of this clone is at base position 88990 of RP11-436C12.

FEATURES

Location/Qualifiers	Source
1..99030	/organism="Homo sapiens"
/db_xref="taxon:9606"	
/chromosome="2"	
/map="2"	
/clone="RP11-260M2"	
/clone_lib="RPC1-11"	
195..483	/rpt_family="MaLR"
484..549	/rpt_family="L1"
550..665	/rpt_family="MaLR"
1489..1530	/rpt_family="(TTG)n"
2391..2427	/rpt_family="AT-rich"
4953..4978	/rpt_family="(TG)n"
4978..5005	/rpt_family="(GA)n"
5211..5320	/rpt_family="L2"
5653..6295	/rpt_family="MER21-group"
7089..7196	/rpt_family="CT-rich"
7939..7986	/rpt_family="MER1-type"
8042..8397	/rpt_family="L1"
8398..8693	/rpt_family="Alu"
8694..9033	/rpt_family="L1"
9045..9232	/rpt_family="Alu"
9356..9611	/rpt_family="Alu"
10944..11018	/rpt_family="(TATATG)n"
11295..11326	/rpt_family="AT-rich"
12027..12521	/note="match to EST A1791679 (NID:g5339395) op98c06.y5"
12079..12639	/note="match to EST AA939008 (NID:g3098921) op98c06.s1"
15316..15343	/rpt_family="AT-rich"
17192..17229	/rpt_family="(TATATG)n"
17230..17258	/rpt_family="(TAGA)n"
18462..18694	/rpt_family="MIR"
18809..18837	

misc_feature /note="assembly_fragment:01062
fragment_chain:1"
2432..5660
/note="assembly_fragment:01573
fragment_chain:1"
5761..8322
/note="assembly_fragment:00193
fragment_chain:1"
8423..13132
/note="assembly_fragment:00386
fragment_chain:1"
13233..20498
/note="assembly_fragment:01407
fragment_chain:2"
20599..25933
/note="assembly_fragment:00948
fragment_chain:2"
26034..31897
/note="assembly_fragment:00950
fragment_chain:3"
31998..43682
/note="assembly_fragment:01201
fragment_chain:2"
43783..55739
/note="assembly_fragment:00950
fragment_chain:3"
55840..62497
/note="assembly_fragment:00010
fragment_chain:3"
62598..66053
/note="assembly_fragment:00516
fragment_chain:3"
66154..71233
/note="assembly_fragment:00799
fragment_chain:4"
71334..78676
/note="assembly_fragment:01288
fragment_chain:4"
78777..83031
/note="assembly_fragment:01254
fragment_chain:5"
83132..87164
/note="assembly_fragment:01420
fragment_chain:5"
87265..96445
/note="assembly_fragment:00385"
96546..99250
/note="assembly_fragment:00401"
99351..109824
/note="assembly_fragment:00662"
109925..114821
/note="assembly_fragment:00921"
114922..117686
/note="assembly_fragment:01024"
117787..123506
/note="assembly_fragment:01232"
123607..142472
/note="assembly_fragment:01449"
142573..151055
/note="assembly_fragment:01538"
151156..153665
/note="assembly_fragment:01612"
29528 c 29581 g 45550 t 2314 others

BASE COUNT 46692 a 29528 c 29581 g 45550 t 2314 others
ORIGIN

Query Match 0.9%; Score 46; DB 78; Length 153665;
Best Local Similarity 100.0%; Pred. No. 5.3e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

45 aggggtagaataattctgtcattgcattcaagactaggacca 90
|||||
DB 151241 AGGGGTAGAAATATTTCTGTCATGGCTCATTCAAAGACTAGGACCA 151196

RESULT 23
AL358892
LOCUS
DEFINITION
5761..8322
/note="assembly_fragment:00193
fragment_chain:1"
8423..13132
/note="assembly_fragment:00386
fragment_chain:1"
13233..20498
/note="assembly_fragment:01407
fragment_chain:2"
20599..25933
/note="assembly_fragment:00948
fragment_chain:2"
26034..31897
/note="assembly_fragment:00950
fragment_chain:3"
31998..43682
/note="assembly_fragment:01201
fragment_chain:2"
43783..55739
/note="assembly_fragment:00950
fragment_chain:3"
55840..62497
/note="assembly_fragment:00010
fragment_chain:3"
62598..66053
/note="assembly_fragment:00516
fragment_chain:3"
66154..71233
/note="assembly_fragment:00799
fragment_chain:4"
71334..78676
/note="assembly_fragment:01288
fragment_chain:4"
78777..83031
/note="assembly_fragment:01254
fragment_chain:5"
83132..87164
/note="assembly_fragment:01420
fragment_chain:5"
87265..96445
/note="assembly_fragment:00385"
96546..99250
/note="assembly_fragment:00401"
99351..109824
/note="assembly_fragment:00662"
109925..114821
/note="assembly_fragment:00921"
114922..117686
/note="assembly_fragment:01024"
117787..123506
/note="assembly_fragment:01232"
123607..142472
/note="assembly_fragment:01449"
142573..151055
/note="assembly_fragment:01538"
151156..153665
/note="assembly_fragment:01612"

AL358892 195380 bp DNA HTG 06-MAR-2001
Mus musculus chromosome 5 clone RP21-417G6, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
AL358892
AL358892.12 GI:11493289
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus
house mouse.
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 195380)
Wall, M.
Direct Submission
Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 30, 2000 this sequence version replaced gi:11414592.

Center: Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseq@har.mrc.ac.uk
----- Project Information
Center project name: dm417G6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 190612 bases at least Q40
Consensus quality: 192224 bases at least Q30
Consensus quality: 193166 bases at least Q20
Insert size: 194680; sum-of-contigs
Insert size: 191947; 3.7% error; agarose-fp
Quality coverage: 7.33x in Q20 bases; sum-of-contigs Quality
coverage: 8.85x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source Location/Qualifiers
1..195380
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP21-417G6"
/clone_lib="RPCI-21"
1..2333
/note="assembly_fragment:02826
fragment_chain:1"
2434..4472
/note="assembly_fragment:00159
fragment_chain:1"
4573..6623
/note="assembly_fragment:00292"
6724..8757
/note="assembly_fragment:02097"
8858..11498
/note="assembly_fragment:02560"
11599..13673
/note="assembly_fragment:03440"
13774..16148
/note="assembly_fragment:03977"
16249..195380
/note="assembly_fragment:04183"
55503 a 40458 c 40664 g 58048 t 707 others
BASE COUNT
ORIGIN


```

repeat_region 33138..33513
/notes="THE1C repeat: matches 1..371 of consensus"
repeat_region 33514..35089
/notes="THE1B-INTERNAL repeat: matches 1..1580 of
consensus"

Query Match
Best Local Similarity 100.0%; Score 47; DB 92; Length 91835;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgagagccccgagtgagcgagtagcgagtcgagtcggcaaccggagg 47
|||||
Db 42873 CCGGAGAGCCCCGAGTGAGCGGAGTAGCGGAGTCGCGCAACCGGAGG 42919

RESULT 21
AL138818
LOCUS AL138818 153665 bp DNA HTG 18-AUG-2000
DEFINITION Homo sapiens chromosome 13 clone RP11-345N9, *** SEQUENCING IN
PROGRESS ***, 24 unordered pieces.
ACCESSION AL138818
VERSION AL138818.5 GI:9863508
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 153665)
Burton, J.
Direct Submission
Submitted (17-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9212026.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA345N9
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 140298 bases at least Q40
Consensus quality: 144322 bases at least Q30
Consensus quality: 147003 bases at least Q20
Insert size: 151365; sum-of-contigs
Insert size: 161736; 2.9% error; agarose-fp
Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
coverage: 3.40x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 2332 2431: contig of 2331 bp in length
* 2432 5660: contig of 3229 bp in length
* 5661 5760: gap of 100 bp
* 5761 8322: contig of 2562 bp in length
* 8323 8422: gap of 100 bp
* 8423 13132: contig of 4710 bp in length
* 13133 13232: gap of 100 bp
* 13233 20498: contig of 7266 bp in length
* 20499 20598: gap of 100 bp
* 20599 25933: contig of 5335 bp in length
* 25934 26033: gap of 100 bp

```

FEATURES

```

source
1..153665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-345N9"
/clone_lib="RPC1-11.2"

misc_feature
1..2331
/notes="assembly fragment:01062
fragment_chain:1"
2432..5660
/notes="assembly fragment:01573
fragment_chain:1"
5761..8322
/notes="assembly fragment:00193
fragment_chain:1"
8423..13132
/notes="assembly fragment:00386
fragment_chain:1"
13233..20498
/notes="assembly fragment:01407
fragment_chain:2"
20599..25933
/notes="assembly fragment:00948
fragment_chain:2"
26034..31897
/notes="assembly fragment:00483
fragment_chain:2"
31998..43682
/notes="assembly fragment:01201
fragment_chain:2"
43783..55739
/notes="assembly fragment:00950
fragment_chain:3"
55840..62497
/notes="assembly fragment:00010
fragment_chain:3"
62598..66053

```


repeat_region /note="L1MB4 repeat: matches 2. .91 of consensus"
817. .1109
repeat_region /note="AluSg repeat: matches 299. .5 of consensus"
1113. .1165
repeat_region /note="L1MA4 repeat: matches 87. .138 of consensus"
1195. .1249
repeat_region /note="AluJ repeat: matches 1. .55 of consensus; incomplete
1252. .1551
repeat_region /note="AluSg repeat: matches 3. .303 of consensus"
1593. .1797
repeat_region /note="MSTB repeat: matches 185. .391 of consensus"
1669. .1796
repeat_region /note="MSTB repeat: matches 289. .422 of consensus"
3346. .3416
repeat_region /note="L1MB4 repeat: matches 859. .929 of consensus"
3418. .3552
repeat_region /note="L1 repeat: matches 4982. .5119 of consensus"
3558. .3938
repeat_region /note="AluSg repeat: matches 1. .282 of consensus;
incomplete repeat"
4035. .4117
repeat_region /note="L1MB3 repeat: matches 5. .76 of consensus"
4120. .4429
repeat_region /note="AluJb repeat: matches 3. .302 of consensus"
4561. .5055
repeat_region /note="L1MB8 repeat: matches 69. .572 of consensus"
5056. .5186
repeat_region /note="FLAM_C repeat: matches 1. .132 of consensus"
5188. .5500
repeat_region /note="L1MB3 repeat: matches 567. .923 of consensus"
6890. .7101
repeat_region /note="MIR repeat: matches 15. .228 of consensus"
7430. .7715
repeat_region /note="AluJb repeat: matches 1. .292 of consensus"
7718. .8028
repeat_region /note="AluX repeat: matches 1. .302 of consensus"
8460. .8763
repeat_region /note="AluSx repeat: matches 302. .1 of consensus"
8979. .9338
repeat_region /note="M1T1A1 repeat: matches 1. .365 of consensus"
9635. .9799
repeat_region /note="L1MA5A repeat: matches 636. .472 of consensus"
9670. .9814
repeat_region /note="L1PA15 repeat: matches 600. .458 of consensus"
11277. .11857
repeat_region /note="L1PA15 repeat: matches 311. .902 of consensus"
12059. .12316
repeat_region /note="THELC repeat: matches 371. .121 of consensus"
12476. .12670
repeat_region /note="AluJo repeat: matches 301. .107 of consensus;
incomplete repeat"
12673. .12959
repeat_region /note="AluSg repeat: matches 286. .1 of consensus"
12968. .13106
repeat_region /note="FLAM_C repeat: matches 133. .2 of consensus"
13148. .13425
repeat_region /note="AluSg repeat: matches 1. .300 of consensus"
13562. .14154
repeat_region /note="L1MB1 repeat: matches 610. .3 of consensus"
14009. .14231
repeat_region /note="L1 repeat: matches 5390. .5167 of consensus"
14258. .14414
repeat_region /note="L1MA4 repeat: matches 742. .905 of consensus"
14483. .14608
repeat_region /note="FLAM_A repeat: matches 1. .126 of consensus"
14620. .14736
repeat_region /note="L1MAA repeat: matches 930. .1042 of consensus"
15223. .15816
repeat_region /note="AluSc repeat: matches 294. .1 of consensus"
15867. .16161
repeat_region /note="AluSx repeat: matches 1. .293 of consensus"
16162. .16364

repeat_region /note="L1 repeat: matches 4313. .4112 of consensus"
16668. .16953
repeat_region /note="AluJ repeat: matches 1. .295 of consensus"
17125. .17538
prim_transcript /note="match: multiple ESTs; low % ID; match: T596663
R48957 H24241 C14861 F03338; match: W88836 N22360 H70851
F02679 N53317; match: R10942 N73317 N20948 R38018 R42572;
match: R62472 R43248 R41198 R38482 R42487; match: R05993
N71787 T64627 F03312"
repeat_region 18303. .18654
repeat_region /note="M1T1A1 repeat: matches 1. .357 of consensus"
18784. .19084
repeat_region /note="AluSx repeat: matches 2. .302 of consensus"
20018. .20312
repeat_region /note="AluSg repeat: matches 298. .2 of consensus"
21304. .21583
repeat_region /note="AluSx repeat: matches 289. .1 of consensus"
22222. .22251
repeat_region /note="15 copies of 2 mer 90 % conserved"
22501. .22555
repeat_region /note="THELC repeat: matches 371. .316 of consensus"
22624. .22950
repeat_region /note="MSTA repeat: matches 372. .1 of consensus"
23311. .23601
repeat_region /note="AluSg repeat: matches 299. .3 of consensus"
23864. .24165
repeat_region /note="AluSp repeat: matches 303. .1 of consensus"
24267. .24562
repeat_region /note="AluSc repeat: matches 4. .299 of consensus"
24559. .24604
unsure /note="single clone"
24931. .25227
repeat_region /note="AluSg repeat: matches 294. .1 of consensus"
25424. .25562
repeat_region /note="AluSg repeat: matches 139. .1 of consensus;
incomplete repeat"
25564. .26134
repeat_region /note="L1MA5 repeat: matches 5390. .5077 of consensus"
25987. .26296
repeat_region /note="L1 repeat: matches 302. .116 of consensus"
26318. .26717
repeat_region /note="MSTA repeat: matches 426. .2 of consensus"
26837. .27018
repeat_region /note="AluSx repeat: matches 302. .116 of consensus;
incomplete repeat"
27019. .27146
repeat_region /note="L1 repeat: matches 5022. .4893 of consensus"
27316. .27534
repeat_region /note="AluJb repeat: matches 302. .85 of consensus;
incomplete repeat"
27720. .27880
repeat_region /note="MSTB repeat: matches 157. .2 of consensus"
28372. .28672
repeat_region /note="AluSp repeat: matches 303. .2 of consensus"
28936. .29047
repeat_region /note="MIR2 repeat: matches 138. .24 of consensus"
29097. .29475
repeat_region /note="M1T2FB repeat: matches 1. .365 of consensus"
29474. .29670
repeat_region /note="MER2 repeat: matches 4. .206 of consensus"
29751. .29797
repeat_region /note="M1T2D repeat: matches 507. .553 of consensus"
30487. .30726
repeat_region /note="AluJo repeat: matches 239. .1 of consensus;
incomplete repeat"
31154. .31451
repeat_region /note="AluSc repeat: matches 298. .1 of consensus"
31628. .32166
repeat_region /note="L1 repeat: matches 4854. .5390 of consensus"
32017. .32906
repeat_region /note="L1PA2 repeat: matches 1. .891 of consensus"
32909. .32938
repeat_region /note="15 copies of 2 mer 87 % conserved"

misc_feature /note="assembly_fragment:03653
fragment_chain:5"
158093..164601
misc_feature /note="assembly_fragment:05568
fragment_chain:5"
164702..195901
misc_feature /note="assembly_fragment:01977
fragment_chain:5"
196002..198807
misc_feature /note="assembly_fragment:00023
fragment_chain:6"
198908..204604
misc_feature /note="assembly_fragment:05254
fragment_chain:6"
204705..207838
misc_feature /note="assembly_fragment:00245
fragment_chain:7"
207939..221335
misc_feature /note="assembly_fragment:03587
fragment_chain:7"
221436..224577
misc_feature /note="assembly_fragment:00642
fragment_chain:8"
224678..228248
misc_feature /note="assembly_fragment:04656
fragment_chain:8"
228349..237923
misc_feature /note="assembly_fragment:01666
fragment_chain:9"
238024..242134
misc_feature /note="assembly_fragment:05529
fragment_chain:9"
242235..250341
misc_feature /note="assembly_fragment:02430
fragment_chain:10"
250442..273524
misc_feature /note="assembly_fragment:00629
fragment_chain:10"
273625..277125
misc_feature /note="assembly_fragment:04232
fragment_chain:11"
277226..282694
misc_feature /note="assembly_fragment:02725
fragment_chain:11"
282795..284888
misc_feature /note="assembly_fragment:04893
fragment_chain:12"
284989..290372
misc_feature /note="assembly_fragment:01882
fragment_chain:12"
290473..298809
misc_feature /note="assembly_fragment:00742"
298910..301400
misc_feature /note="assembly_fragment:00762"
301501..303502
misc_feature /note="assembly_fragment:01554"
303603..306492
misc_feature /note="assembly_fragment:02456"
306593..309852
misc_feature /note="assembly_fragment:03413"
309953..312398
misc_feature /note="assembly_fragment:03967"
312499..315866
misc_feature /note="assembly_fragment:04102"
315967..318753
misc_feature /note="assembly_fragment:04448"
318854..323190
misc_feature /note="assembly_fragment:04616"
323291..329923
misc_feature /note="assembly_fragment:04963"
330024..332405
misc_feature /note="assembly_fragment:06188"
332506..335631

misc_feature /note="assembly_fragment:06207"
335732..338304
misc_feature /note="assembly_fragment:06594"
338405..340969
BASE COUNT 100857 a 71110 c 68176 g 96314 t 4512 others
ORIGIN

Query Match 1.1%; Score 56; DB 81; Length 340969;
Best Local Similarity 100.0%; Pred. No. 2.3e-18;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3006 gaaaattattgtcttctaccagagatatttccatatacaattcacttttgaca 3065
|||||
Db 61173 GAAAAATTATTTGCTCTTCTACAGAGATATTGTTCCATACAAATTCACCTTTTGCCA 61114
|||||

QY 3066 catgcccagattatgtcaaggtacagagatattgaacaacttaagatgt 3115
|||||
Db 61113 CATGACCAGATTATGTCAAAGTACAGGATATTGAACAACCTTAAGATGT 61064
|||||

RESULT 20
HS26H23
LOCUS
DEFINITION HS26H23 91835 bp DNA PRI 22-NOV-1999
Human DNA sequence from PAC 26H23, BRCA2 gene region chromosome
13q12-13 contains ESTs, CpG island.
ACCESSION 284467 285990 285991 285992 285993
VERSION 284467.1 GI:2104578
KEYWORDS 13q12-13; CpG island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 91835)
AUTHORS Williamson, H.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 17, 1997 this sequence version replaced gi:1806009.
IMPORTANT:
This sequence is not the entire insert of clone 26H23. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 26H23 is at 1 in this sequence. The true
right end of clone 130N4 is at 34585.
The true left end of clone 267P19 is at 91732.
26H23 is from the human PAC library described in Ioannou A.P. et al
Nature Genet 6, 84-89.

FEATURES
source
1. 91835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-13"
/clone_lib="RPC1-1"
/clone="XX-26H23"
1. 83
/note="AluSc repeat: matches 84. .1 of consensus;
incomplete repeat"
94. .396
repeat_region /note="AluSx repeat: matches 302. .1 of consensus"
451. .710
repeat_region /note="L1 repeat: matches 4886. .5146 of consensus"
708. .815
repeat_region

PCR Profile:

Presoak: 94 degrees C for 4.00 minutes
 Denaturation: 94 degrees C for 50.0 seconds
 Annealing: 58 degrees C for 1.50 minutes
 Polymerization: 72 degrees C for 1.00 minutes
 PCR Cycles: 30
 Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:

Template: 10 ng
 Primer: each 5 pM
 dNTPs: 4 mM
 Taq Polymerase: 0.5 U
 Total Vol: 20 uL

Buffer:

Mg2+: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 Gelatin: .001 %

FEATURES

source
 Location/Qualifiers
 1..150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="24.00 cR from top of Chr13 linkage group"
 /clone_lib="Human Thudson EST"
 /note="STS derived from sequences in dbEST and the
 Unigene collection."
 STS
 primer_bind
 1..150
 primer_bind
 1..20
 complement(127..150)
 BASE COUNT 43 a 26 c 38 g 42 t 1 others
 ORIGIN

Query Match 1.9%; Score 99; DB 54; Length 150;

Best Local Similarity 99.3%; Pred. No. 6.2e-42;
 Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4491 cacaaatgggactgtgaagagtgacagtggtgaccttacttggtagcccccatacatt 4550
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1 CACAAATGGGACTGCTGAAGAGTGACAGTGGACCTTACTTTGGTAGCCCCATACATT 60
 QY 4551 tgggtcacatgttttagccatcacatggttaacattgactatggaggtcttggaaagtg 4610
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 61 TGTGGTCACATGCTTTAGGCATACNCATGCTAACATTGACTGGAGTCTGTGAAAGTG 120
 QY 4611 taatgtcgtatggtctatgtagacataaaga 4640
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 121 TAATGTGGATGGCTATGTAGACATAAAGA 150

RESULT 19

AL512630/c
 LOCUS 340969 bp DNA HTG 21-MAR-2001
 DEFINITION Mus musculus chromosome 5 clone RP21-583E8, *** SEQUENCING IN
 PROGRESS ***, in unordered pieces.
 ACCESSION AL512630
 VERSION AL512630.2 GI:13443466
 KEYWORDS HTG; HTGS, PHASE1.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 340969)
 AUTHORS Sims, S.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonequests@sanger.ac.uk
 On Mar 24, 2001 this sequence version replaced gi:12193249.
 COMMENT
 ----- Genome Center
 Center: UK Medical Research Council
 Center code: UK-MRC

Web site: http://mrcseq.har.mrc.ac.uk
 Contact: mouse@har.mrc.ac.uk
 ----- Project Information
 Center project name: dm583E8
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 323281 bases at least Q40
 Consensus quality: 329435 bases at least Q30
 Consensus quality: 333234 bases at least Q20
 Insert size: 336469; sum-of-contigs
 Insert size: 189900; 1.3% error; agarose-fp
 Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality
 coverage: 11.68x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source
 Location/Qualifiers
 1..340969
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="5"
 /clone_lib="RP21-583E8"
 /clone_lib="RPGI-21"
 1..10854
 /note="assembly_fragment:06116
 fragment_chain:1"
 misc_feature
 10955..16884
 /note="assembly_fragment:05022
 fragment_chain:1"
 misc_feature
 16985..22344
 /note="assembly_fragment:04089
 fragment_chain:1"
 misc_feature
 22445..36631
 /note="assembly_fragment:04629
 fragment_chain:1"
 misc_feature
 36732..63941
 /note="assembly_fragment:02431
 fragment_chain:1"
 misc_feature
 64042..76959
 /note="assembly_fragment:06652
 fragment_chain:2"
 misc_feature
 77060..80361
 /note="assembly_fragment:02856
 fragment_chain:2"
 misc_feature
 80462..84581
 /note="assembly_fragment:05282
 fragment_chain:2"
 misc_feature
 84682..89169
 /note="assembly_fragment:01199
 fragment_chain:2"
 misc_feature
 89270..94454
 /note="assembly_fragment:00854
 fragment_chain:3"
 misc_feature
 94555..99344
 /note="assembly_fragment:04769
 fragment_chain:3"
 misc_feature
 99445..105364
 /note="assembly_fragment:03592
 fragment_chain:3"
 misc_feature
 105465..112857
 /note="assembly_fragment:02797
 fragment_chain:4"
 misc_feature
 112958..116353
 /note="assembly_fragment:05253
 fragment_chain:4"
 misc_feature
 116454..119029
 /note="assembly_fragment:01663
 fragment_chain:4"
 misc_feature
 119130..157992


```

* 51027 51126: gap of 100 bp
* 51127 51836: contig of 710 bp in length
* 51837 51936: gap of 100 bp
* 51937 52671: contig of 735 bp in length
* 52672 52771: gap of 100 bp
* 52772 53502: contig of 731 bp in length
* 53503 53602: gap of 100 bp
* 53603 54343: contig of 741 bp in length
* 54344 54443: gap of 100 bp
* 54444 55141: contig of 698 bp in length
* 55142 55241: gap of 100 bp
* 55242 55950: contig of 709 bp in length
* 55951 56050: gap of 100 bp

Query Match 1.9%; Score 100; DB 77; Length 69740;
Best Local Similarity 100.0%; Pred. No. 3.2e-42;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 aagttgggaaccccaacagcctacaaatccttcttctggaatgatcaagtttctcttg 1880
|||||
Db 13983 AAGTTGGCAACCCCAACAGCCTACAAATCCTTCTCTGGAATGATCAAGTTCTCTTG 14042

QY 1881 gagaggatagcactgtgcacatagataccgaatctatca 1920
|||||
Db 14043 GAGAGGATAGCACCCTGTGCACATAGATACCGAATCTATCA 14082

RESULT 17
G43328 150 bp mRNA STS 27-JAN-1999
LOCUS
DEFINITION WIAF-1982-STS Human Thudson EST Homo sapiens STS cDNA, sequence
tagged site.
ACCESSION G43328
VERSION G43328.1 GI:4192245
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150)
AUTHORS Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
JOURNAL MEDLINE
COMMENT 98248615
Synonyms: EST397175a, EST397175
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CACAAATGGACTGCTGAA
Primer B: TCTTTATGTCTACATAGCATCGC
STS size: 150
PCR Profile:
Presoak: 94 degrees C for 4.00 minutes
Denaturation: 94 degrees C for 50.0 seconds
Annealing: 58 degrees C for 1.50 minutes
Polymerization: 72 degrees C for 1.00 minutes
PCR Cycles: 30
Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: 4 nM
Taq Polymerase: 0.5 U

```

```

Total Vol: 20 uL
Buffer:
Mg2+: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Gelatin: .001 %.
Location/Qualifiers
1..150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="24.00 cR from top of Chr13 linkage group"
/clone_lib="Human Thudson EST"
/note="STSs derived from sequences in dbEST and the
Unigene collection."
STS
primer_bind 1..150
primer_bind 1..20
BASE COUNT 43 a 26 c 38 g 42 t 1 others
ORIGIN

Query Match 1.9%; Score 99; DB 54; Length 150;
Best Local Similarity 99.3%; Pred. No. 6.2e-42;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4491 cacaaaatgggactgtgaagagtgacagttggacattgaccttacttgggtgacccatacatt 4550
|||||
Db 1 CACAAAATGGGACTGCTGAAGAGTGGACAGTGGACCTTACTTTGGTGACCCCATACNTT 60

QY 4551 tgtgtcacatgcttttagccatacacatggttaacattgacctatggagctctgtgaaagt 4610
|||||
Db 61 TGTGTCACATGCTTTAGCCATACACATGGTAAACATTTGACTATGAGTCTTGTGAAGTG 120

QY 4611 taatgtgcgatggctatgtagacataaaga 4640
|||||
Db 121 TAATGTGCGATGGCTATGTAGACATAAAGA 150

RESULT 18
G43329 150 bp mRNA STS 27-JAN-1999
LOCUS
DEFINITION WIAF-1983-STS Human Thudson EST Homo sapiens STS cDNA, sequence
tagged site.
ACCESSION G43329
VERSION G43329.1 GI:4192246
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150)
AUTHORS Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
JOURNAL MEDLINE
COMMENT 98248615
Synonyms: EST397175b, EST397175
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CACAAATGGACTGCTGAA
Primer B: TCTTTATGTCTACATAGCATCGC
STS size: 150

```


<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L12085
 Center clone name: 179_J_16

* NOTE: This record contains 86 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 737: contig of 737 bp in length
 * 738 837: gap of 100 bp
 * 838 1551: contig of 714 bp in length
 * 1552 1651: gap of 100 bp
 * 1652 2343: contig of 692 bp in length
 * 2344 2443: gap of 100 bp
 * 2444 3138: contig of 695 bp in length
 * 3139 3238: gap of 100 bp
 * 3239 3953: contig of 715 bp in length
 * 3954 4053: gap of 100 bp
 * 4054 4722: contig of 669 bp in length
 * 4723 4822: gap of 100 bp
 * 4823 5530: contig of 708 bp in length
 * 5531 5630: gap of 100 bp
 * 5631 6330: contig of 700 bp in length
 * 6331 6430: gap of 100 bp
 * 6431 7165: contig of 735 bp in length
 * 7166 7265: gap of 100 bp
 * 7266 7957: contig of 692 bp in length
 * 7958 8057: gap of 100 bp
 * 8058 8788: contig of 731 bp in length
 * 8789 8888: gap of 100 bp
 * 8889 9619: contig of 731 bp in length
 * 9620 9719: gap of 100 bp
 * 9720 10467: contig of 748 bp in length
 * 10468 10567: gap of 100 bp
 * 10568 11265: contig of 698 bp in length
 * 11266 11365: gap of 100 bp
 * 11366 12055: contig of 690 bp in length
 * 12056 12155: gap of 100 bp
 * 12156 12847: contig of 692 bp in length
 * 12848 12947: gap of 100 bp
 * 12948 13662: contig of 715 bp in length
 * 13663 13762: gap of 100 bp
 * 13763 14476: contig of 714 bp in length
 * 14477 14576: gap of 100 bp
 * 14577 15302: contig of 726 bp in length
 * 15303 15402: gap of 100 bp
 * 15403 16105: contig of 703 bp in length
 * 16106 16205: gap of 100 bp
 * 16206 16942: contig of 737 bp in length
 * 16943 17042: gap of 100 bp
 * 17043 17786: contig of 744 bp in length
 * 17787 17886: gap of 100 bp
 * 17887 18612: contig of 726 bp in length
 * 18613 18712: gap of 100 bp
 * 18713 19420: contig of 708 bp in length
 * 19421 19520: gap of 100 bp
 * 19521 20216: contig of 696 bp in length
 * 20217 20316: gap of 100 bp
 * 20317 21022: contig of 706 bp in length
 * 21023 21122: gap of 100 bp

* 21123 21827: contig of 705 bp in length
 * 21828 21927: gap of 100 bp
 * 21928 22639: contig of 712 bp in length
 * 22640 22739: gap of 100 bp
 * 22740 23453: contig of 714 bp in length
 * 23454 23553: gap of 100 bp
 * 23554 24279: contig of 726 bp in length
 * 24280 24379: gap of 100 bp
 * 24380 25082: contig of 703 bp in length
 * 25083 25182: gap of 100 bp
 * 25183 25907: contig of 725 bp in length
 * 25908 26007: gap of 100 bp
 * 26008 26732: contig of 725 bp in length
 * 26733 26832: gap of 100 bp
 * 26833 27561: contig of 729 bp in length
 * 27562 27661: gap of 100 bp
 * 27662 28358: contig of 697 bp in length
 * 28359 28458: gap of 100 bp
 * 28459 29136: contig of 678 bp in length
 * 29137 29236: gap of 100 bp
 * 29237 29936: contig of 700 bp in length
 * 29937 30036: gap of 100 bp
 * 30037 30736: contig of 700 bp in length
 * 30737 30836: gap of 100 bp
 * 30837 31561: contig of 725 bp in length
 * 31562 31661: gap of 100 bp
 * 31662 32380: contig of 719 bp in length
 * 32381 32480: gap of 100 bp
 * 32481 33224: contig of 744 bp in length
 * 33225 33324: gap of 100 bp
 * 33325 34053: contig of 729 bp in length
 * 34054 34153: gap of 100 bp
 * 34154 34877: contig of 724 bp in length
 * 34878 34977: gap of 100 bp
 * 34978 35703: contig of 726 bp in length
 * 35704 35803: gap of 100 bp
 * 35804 36503: contig of 700 bp in length
 * 36504 36603: gap of 100 bp
 * 36604 37315: contig of 712 bp in length
 * 37316 37415: gap of 100 bp
 * 37416 38120: contig of 705 bp in length
 * 38121 38220: gap of 100 bp
 * 38221 38916: contig of 696 bp in length
 * 38917 39016: gap of 100 bp
 * 39017 39712: contig of 696 bp in length
 * 39713 39812: gap of 100 bp
 * 39813 40535: contig of 723 bp in length
 * 40536 40635: gap of 100 bp
 * 40636 41335: contig of 700 bp in length
 * 41336 41435: gap of 100 bp
 * 41436 42087: contig of 652 bp in length
 * 42088 42187: gap of 100 bp
 * 42188 42914: contig of 727 bp in length
 * 42915 43014: gap of 100 bp
 * 43015 43738: contig of 724 bp in length
 * 43739 43838: gap of 100 bp
 * 43839 44553: contig of 715 bp in length
 * 44554 44653: gap of 100 bp
 * 44654 45353: contig of 700 bp in length
 * 45354 45453: gap of 100 bp
 * 45454 46147: contig of 694 bp in length
 * 46148 46247: gap of 100 bp
 * 46248 46939: contig of 692 bp in length
 * 46940 47039: gap of 100 bp
 * 47040 47730: contig of 691 bp in length
 * 47731 47830: gap of 100 bp
 * 47831 48537: contig of 707 bp in length
 * 48538 48637: gap of 100 bp
 * 48638 49354: contig of 717 bp in length
 * 49355 49454: gap of 100 bp
 * 49455 50194: contig of 740 bp in length
 * 50195 50294: gap of 100 bp
 * 50295 51026: contig of 732 bp in length

* 20217 20316: gap of 100 bp
 * 20317 21022: contig of 706 bp in length
 * 21023 21122: gap of 100 bp
 * 21123 21827: contig of 705 bp in length
 * 21828 21927: gap of 100 bp
 * 21928 22639: contig of 712 bp in length
 * 22640 22739: gap of 100 bp
 * 22740 23453: contig of 714 bp in length
 * 23454 23553: gap of 100 bp
 * 23554 24279: contig of 726 bp in length
 * 24280 24379: gap of 100 bp
 * 24380 25082: contig of 703 bp in length
 * 25083 25182: gap of 100 bp
 * 25183 25907: contig of 725 bp in length
 * 25908 26007: gap of 100 bp
 * 26008 26732: contig of 725 bp in length
 * 26733 26832: gap of 100 bp
 * 26833 27561: contig of 729 bp in length
 * 27562 27661: gap of 100 bp
 * 27662 28358: contig of 697 bp in length
 * 28359 28458: gap of 100 bp
 * 28459 29136: contig of 678 bp in length
 * 29137 29236: gap of 100 bp
 * 29237 29936: contig of 700 bp in length
 * 29937 30036: gap of 100 bp
 * 30037 30736: contig of 700 bp in length
 * 30737 30836: gap of 100 bp
 * 30837 31561: contig of 725 bp in length
 * 31562 31661: gap of 100 bp
 * 31662 32380: contig of 719 bp in length
 * 32381 32480: gap of 100 bp
 * 32481 33224: contig of 744 bp in length
 * 33225 33324: gap of 100 bp
 * 33225 34053: contig of 729 bp in length
 * 34054 34153: gap of 100 bp
 * 34154 34877: contig of 724 bp in length
 * 34878 34977: gap of 100 bp
 * 34978 35703: contig of 726 bp in length
 * 35704 35803: gap of 100 bp
 * 35804 36503: contig of 700 bp in length
 * 36504 36603: gap of 100 bp
 * 36604 37315: contig of 712 bp in length
 * 37316 37415: gap of 100 bp
 * 37416 38120: contig of 705 bp in length
 * 38121 38220: gap of 100 bp
 * 38221 38916: contig of 696 bp in length
 * 38917 39016: gap of 100 bp
 * 39017 39712: contig of 696 bp in length
 * 39713 39812: gap of 100 bp
 * 39813 40535: contig of 723 bp in length
 * 40536 40635: gap of 100 bp
 * 40636 41335: contig of 700 bp in length
 * 41336 41435: gap of 100 bp
 * 41436 42087: contig of 652 bp in length
 * 42088 42187: gap of 100 bp
 * 42188 42914: contig of 727 bp in length
 * 42915 43014: gap of 100 bp
 * 43015 43738: contig of 724 bp in length
 * 43739 43838: gap of 100 bp
 * 43839 44533: contig of 715 bp in length
 * 44534 44653: gap of 100 bp
 * 44654 45353: contig of 700 bp in length
 * 45354 45453: gap of 100 bp
 * 45454 46147: contig of 694 bp in length
 * 46148 46247: gap of 100 bp
 * 46248 46939: contig of 692 bp in length
 * 46940 47039: gap of 100 bp
 * 47040 47730: contig of 691 bp in length
 * 47731 47830: gap of 100 bp
 * 47831 48537: contig of 707 bp in length
 * 48538 48637: gap of 100 bp
 * 48638 49354: contig of 717 bp in length
 * 49355 49454: gap of 100 bp

* 49455 50194: contig of 740 bp in length
 * 50195 50294: gap of 100 bp
 * 50295 51026: contig of 732 bp in length
 * 51027 51126: gap of 100 bp
 * 51127 51836: contig of 710 bp in length
 * 51837 51936: gap of 100 bp
 * 51937 52671: contig of 735 bp in length
 * 52672 52771: gap of 100 bp
 * 52772 53502: contig of 731 bp in length
 * 53503 53602: gap of 100 bp
 * 53603 54343: contig of 741 bp in length
 * 54344 54443: gap of 100 bp
 * 54444 55141: contig of 698 bp in length
 * 55142 55241: gap of 100 bp
 * 55242 55950: contig of 709 bp in length
 * 55951 56050: gap of 100 bp

Query Match 2.3%; Score 120; DB 77; Length 69740;
 Best Local Similarity 100.0%; Pred. No. 4.9e-53;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 agagcaatgataatgagcgagcgctcacaagtgtttaaactactggcaaaatgtttgggg 969
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 28043 AGACCAATGATAATGAGGAGCGCCTACAAAGTTGTAAACTACTGGCAAAATGTTGGG 27984
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 970 caaagattcagaattgcttctcaaaacagccacttggcagtcgacttggcaggt 1029
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 27983 CAAAGGATTCAGAATTGGCTTCTCAAAACAGCCACTTTGGCAGTGTACTTTGGCAGGT 27924

RESULT 16

AC087748 69740 bp DNA HTG 22-JAN-2001
 Homo sapiens chromosome 18 clone CTB-179J16 map 18, LOW-PASS
 LOCUS AC087748
 DEFINITION AC087748
 ACCESSION AC087748
 VERSION AC087748.1 GI:12331536
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 69740)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome 18, clone CTB-179J16
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 69740)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A.,
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Gallegher, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Meldrum, J., Meneus, L., Mihova, I., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
 Zemdek, L., Zimmer, A. and Zody, M.
 Direct Submission
 TITLE
 JOURNAL
 COMMENT Submitted (22-JAN-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)


```
/note="assembly_fragment:00874
fragment_chain:2"
misc_feature 147136..149710
/note="assembly_fragment:00391
fragment_chain:3"
misc_feature 149811..163642
/note="assembly_fragment:00398
fragment_chain:3"
BASE COUNT 54756 a 31368 c 29374 g 47337 t 807 others
ORIGIN

Query Match 3.0%; Score 156; DB 79; Length 163642;
Best Local Similarity 99.5%; Pred. No. 1.8e-72;
Matches 206; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 172 agatgtgtgaaactttatgatatgacacgactcgaagaaagagacgttt 231
|||||
Db 32938 AGATGTTGTGAACACTTTATGGATATGACACGACCTCTGAAGAAGAAAGAGCTTT 32879
|||||

QY 232 atttaaacctagctttacatcttgccttcagattttttctcgaagcatcctggtaaagatg 291
|||||
Db 32878 ATTAAACCTAGCTTTACATCTTGCCTTCAGATTTTCTCAAGCATCTCGATAAAGATG 32819
|||||

QY 292 ttgcgttaactgtagcctgctgcttgcgtatatttttcaggatttatgctcctgaagctc 351
|||||
Db 32818 TTGCGTTACTGGTAGCTGCTGCTGCTGATATTTTCAGGATTTATGCTCTCGAAGCTC 32759
|||||

QY 352 cttacacatccctgataaactaaagg 378
|||||
Db 32758 CTACACATCCCTGATAAAGG 32732
|||||

RESULT 15
AC087748/c
LOCUS AC087748 69740 bp DNA HTG 22-JAN-2001
DEFINITION Homo sapiens chromosome 18 clone CTB-179J16 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087748
VERSION AC087748.1 GI:12331536
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 69740)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone CTB-179J16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 69740)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouckhalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Plerre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
JOURNAL Submitted (22-JAN-2001) Whitehead Institute/MIT Center for Genome
```

Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12085
Center clone name: 179_J16

* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 737: contig of 737 bp in length
* 738 837: gap of 100 bp
* 838 1551: contig of 714 bp in length
* 1552 1651: gap of 100 bp
* 1652 2343: contig of 692 bp in length
* 2344 2443: gap of 100 bp
* 2444 3138: contig of 695 bp in length
* 3139 3238: gap of 100 bp
* 3239 3953: contig of 715 bp in length
* 3954 4053: gap of 100 bp
* 4054 4722: contig of 669 bp in length
* 4723 4822: gap of 100 bp
* 4823 5530: contig of 708 bp in length
* 5531 5630: gap of 100 bp
* 5631 6330: contig of 700 bp in length
* 6331 6430: gap of 100 bp
* 6431 7165: contig of 735 bp in length
* 7166 7265: gap of 100 bp
* 7266 7957: contig of 692 bp in length
* 7958 8057: gap of 100 bp
* 8058 8788: contig of 731 bp in length
* 8789 8888: gap of 100 bp
* 8889 9619: contig of 731 bp in length
* 9620 9719: gap of 100 bp
* 9720 10467: contig of 748 bp in length
* 10468 10567: gap of 100 bp
* 10568 11265: contig of 698 bp in length
* 11266 11365: gap of 100 bp
* 11366 12055: contig of 690 bp in length
* 12056 12155: gap of 100 bp
* 12156 12847: contig of 692 bp in length
* 12848 12947: gap of 100 bp
* 12948 13662: contig of 715 bp in length
* 13663 13762: gap of 100 bp
* 13763 14476: contig of 714 bp in length
* 14477 14576: gap of 100 bp
* 14577 15302: contig of 726 bp in length
* 15303 15402: gap of 100 bp
* 15403 16105: contig of 703 bp in length
* 16106 16205: gap of 100 bp
* 16206 16942: contig of 737 bp in length
* 16943 17042: gap of 100 bp
* 17043 17786: contig of 744 bp in length
* 17787 17886: gap of 100 bp
* 17887 18612: contig of 726 bp in length
* 18613 18712: gap of 100 bp
* 18713 19420: contig of 708 bp in length
* 19421 19520: gap of 100 bp
* 19521 20216: contig of 696 bp in length


```
repeat_region 46910..46988
/note="L1 element fragment"
repeat_region 47081..47119
/note="L1 element fragment"
repeat_region 47168..47431
/note="L1 element fragment"
repeat_region 47615..47902
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 47922..48213
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 49426..49717
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 50248..50533
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 51097..51380
/note="Alu repeat: matches 1. .304 of consensus"
repeat_region 62318..62543
/partial
/note="Alu repeat: matches 302. .66 of consensus"
repeat_region 63358..63613
/partial
/note="Alu repeat: matches 306. .34 of consensus"
repeat_region 66191..66236
/note="23 copies of 2 mer 80 % conserved"
repeat_region 68771..69062
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 70282..70559
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 70590..70881
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 75593..75883
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 75974..76251
/partial
/note="Alu repeat: matches 294. .1 of consensus"
repeat_region 76551..76608
/note="MIR element fragment"
repeat_region 78482..78832
/note="MER11B element fragment"
repeat_region 78574..78743
/note="MER11A element fragment"
repeat_region 78941..79021
/note="MER11B element fragment"
repeat_region 79249..79709
/note="MER11A element fragment"
repeat_region 80233..80302
/note="SVA element fragment"
repeat_region 80355..80535
/partial
/note="Alu repeat: matches 213. .20 of consensus"
repeat_region 80563..80619
/partial
/note="Alu repeat: matches 277. .221 of consensus"
repeat_region 80651..82294
/note="SVA element fragment"
repeat_region 82753..83043
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region 83149..83296
/partial
```

Query Match 3.1%; Score 163; DB 92; Length 113704;
Best Local Similarity 100.0%; Pred. No. 2.9e-76;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2026 aggtactctattacacccattctattcttctgctgaacatttgatcattac 2085
|||||
DB 74538 AGGTACTCTCATTATACATCCCTCATTCATTCCTGCTGAAACATTTGAATCATTC 74597

QY 2086 tggctgtctgaaatggatgataaaagtagcagaagctgactacacaaattttcaaaa 2145
|||||
DB 74598 TGGCTGTCTGAAATGGATGATGAAAAAGTAGCAGAGCTGACTACAAATTTTCAAAA 74657
|||||

QY 2146 acacaggaagcaaaatgaagaggatttccacacatcagatc 2188
|||||
DB 74658 ACACAGGAAGCAAAATGAAGAGGATTTTCCACACATCATGATC 74700
|||||

RESULT 13
ALJ53724
LOCUS
DEFINITION
Homo sapiens chromosome 13 clone RP11-448I13, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION
ALJ53724
VERSION
ALJ53724.3 GI:9930876
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 163642)
Burton,J.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926534.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA448I13
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159306 bases at least Q40
Consensus quality: 161143 bases at least Q30
Consensus quality: 162084 bases at least Q20
Insert size: 162842; sum-of-contigs
Insert size: 163676; 2.5% error; agarose-fp
Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality
coverage: 4.42x in Q20 bases; agarose-fp

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20261: contig of 20261 bp in length
* 20262 20361: gap of 100 bp
* 20362 38140: contig of 17779 bp in length
* 38141 38240: gap of 100 bp
* 38241 85018: contig of 46778 bp in length
* 85019 85118: gap of 100 bp
* 85119 93385: contig of 8267 bp in length
* 93386 93485: gap of 100 bp
* 93486 96470: contig of 2985 bp in length
* 96471 96570: gap of 100 bp
* 96571 137301: contig of 40731 bp in length
* 137302 137401: gap of 100 bp
* 137402 147035: contig of 9634 bp in length
* 147036 147135: gap of 100 bp
* 147136 149710: contig of 2575 bp in length
* 149711 149810: gap of 100 bp
* 149811 163842: contig of 13832 bp in length.

```

QY 4255 cattcaaaccaatttcaaaattatttgcacaaagtctctaaatttgcataacatacatatt 4314
|||||
DB 34062 CATTTCACCAAAATTTCAATATTTTGCACAAAGTTCTTAAATTTGTAACATACATATT 34121
|||||

QY 4315 gctgtatttaaacctcatatatttagccattacacatagtagta 4357
|||||
DB 34122 GCTGTATTAAATTCATATATTATTAGCCCATACACTAGTAGTA 34164
|||||

RESULT 12
HS267P19 113704 bp DNA PRI 22-NOV-1999
LOCUS
DEFINITION Human DNA sequence from cosmid 267P19, BRCA2 gene region chromosome
13q12-13 contains polymorphic CA repeat.
ACCESSION 275889
VERSION 275889
KEYWORDS 13q12-13; repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 113704)
AUTHORS Odell,C.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-1996) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, UK. E-mail enquires: humquery@sanger.ac.uk
requests: clonesrequest@sanger.ac.uk
COMMENT
IMPORTANT:
This sequence is not the entire insert of clone 267P19. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 267P19 is at 1 in this sequence. The
true right end of clone 26H23 is at 24539.
The true left end of clone 49J10 is at 113606.
267P19 is from the human PAC library.
FEATURES
source
1..113704
/organism="Homo sapiens"
/db_xref="taxon.9606"
/chromosome="13"
/map="13q12-13"
/clone="XX-267P19"
repeat_region
1..235
/partial
/notes="Alu repeat: matches 252. .1 of consensus"
370..662
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
4849..5140
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
5367..5660
/notes="Alu repeat: matches 1. .308 of consensus"
6080..6187
/notes="2 copies of 54 mer 87 & conserved"
8133..8401
/notes="MER7A element fragment"
8623..8748
/notes="L1 element fragment"
9617..9711
/partial
/notes="Alu repeat: matches 305. .210 of consensus"
9724..9876
/partial
/notes="Alu repeat: matches 150. .1 of consensus"
10567..10594

```

```

repeat_region
10616..10900
/partial
/notes="Alu repeat: matches 300. .1 of consensus"
10960..10985
/notes="13 copies of 2 mer 92 & conserved"
11912..12100
/partial
/notes="Alu repeat: matches 268. .65 of consensus"
17050..17359
/notes="Alu repeat: matches 1. .306 of consensus"
17639..17888
/partial
/notes="Alu repeat: matches 34. .306 of consensus"
19151..19980
/notes="L1 element fragment"
19981..20264
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
20298..21190
/notes="L1 element fragment"
21199..21488
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
21538..22226
/notes="L1 element fragment"
25696..25989
/notes="Alu repeat: matches 1. .308 of consensus"
26902..27167
/notes="2 copies of 133 mer 93 & conserved"
27850..27930
/notes="L1 element fragment"
28002..28043
/notes="L1 element fragment"
28323..28780
/notes="L1 element fragment"
29176..29449
/notes="Alu repeat: matches 1. .306 of consensus"
29904..30192
/notes="Alu repeat: matches 1. .308 of consensus"
30354..30663
/notes="L1 element fragment"
30683..30972
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
30977..31204
/notes="L1 element fragment"
34819..35121
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
37059..37282
/notes="MLT2B2 element fragment"
37453..37506
/notes="MLT2B2 element fragment"
38414..38706
/partial
/notes="Alu repeat: matches 308. .1 of consensus"
39516..39563
/notes="24 copies of 2 mer 81 & conserved"
41244..41535
/notes="Alu repeat: matches 1. .308 of consensus"
43994..44242
/partial
/notes="Alu repeat: matches 44. .308 of consensus"
complement(44369..44677)
/notes="match: 217151 DNA segment containing (CA) repeat"
complement(44479..44510)
/notes="16 copies of CA 100% conserved; Polymorphic by
comparison with 217151"
44702..44889
/notes="2 copies of 94 mer 85 & conserved"
45935..46229
/notes="Alu repeat: matches 1. .308 of consensus"

```


AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campolano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Donino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Gatagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczek, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivier, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10161
Center clone name: 660_R_19

* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 685: contig of 685 bp in length
* 686 785: gap of 100 bp
* 786 1513: contig of 728 bp in length
* 1514 1613: gap of 100 bp
* 1614 2354: contig of 741 bp in length
* 2355 2454: gap of 100 bp
* 2455 3184: contig of 730 bp in length
* 3185 3284: gap of 100 bp
* 3285 3959: contig of 675 bp in length
* 3960 4059: gap of 100 bp
* 4060 4746: contig of 687 bp in length
* 4747 4846: gap of 100 bp
* 4847 5564: contig of 718 bp in length
* 5565 5664: gap of 100 bp
* 5665 6360: contig of 696 bp in length
* 6361 6460: gap of 100 bp
* 6461 7132: contig of 672 bp in length
* 7133 7232: gap of 100 bp
* 7233 7937: contig of 705 bp in length
* 7938 8037: gap of 100 bp
* 8038 8749: contig of 712 bp in length
* 8750 8849: gap of 100 bp
* 8850 9522: contig of 673 bp in length
* 9523 9622: gap of 100 bp
* 9623 10356: contig of 734 bp in length
* 10357 10456: gap of 100 bp

* 10457 11181: contig of 725 bp in length
* 11182 11281: gap of 100 bp
* 11282 11975: contig of 694 bp in length
* 11976 12075: gap of 100 bp
* 12076 12785: contig of 710 bp in length
* 12786 12885: gap of 100 bp
* 12886 13557: contig of 672 bp in length
* 13558 13657: gap of 100 bp
* 13658 14347: contig of 690 bp in length
* 14348 14447: gap of 100 bp
* 14448 15132: contig of 685 bp in length
* 15133 15232: gap of 100 bp
* 15233 15912: contig of 680 bp in length
* 15913 16012: gap of 100 bp
* 16013 16731: contig of 719 bp in length
* 16732 16831: gap of 100 bp
* 16832 17506: contig of 675 bp in length
* 17507 17606: gap of 100 bp
* 17607 18339: contig of 733 bp in length
* 18340 18439: gap of 100 bp
* 18440 19109: contig of 670 bp in length
* 19110 19209: gap of 100 bp
* 19210 19900: contig of 691 bp in length
* 19901 20000: gap of 100 bp
* 20001 20693: contig of 693 bp in length
* 20694 20793: gap of 100 bp
* 20794 21464: contig of 671 bp in length
* 21465 21564: gap of 100 bp
* 21565 22245: contig of 681 bp in length
* 22246 22345: gap of 100 bp
* 22346 23049: contig of 704 bp in length
* 23050 23149: gap of 100 bp
* 23150 23857: contig of 708 bp in length
* 23858 23957: gap of 100 bp
* 23958 24685: contig of 728 bp in length
* 24686 24785: gap of 100 bp
* 24786 25499: contig of 714 bp in length
* 25500 25599: gap of 100 bp
* 25600 26302: contig of 703 bp in length
* 26303 26402: gap of 100 bp
* 26403 27108: contig of 706 bp in length
* 27109 27208: gap of 100 bp
* 27209 27910: contig of 702 bp in length
* 27911 28010: gap of 100 bp
* 28011 28720: contig of 710 bp in length
* 28721 28820: gap of 100 bp
* 28821 29484: contig of 664 bp in length
* 29485 29584: gap of 100 bp
* 29585 30328: contig of 744 bp in length
* 30329 30428: gap of 100 bp
* 30429 31111: contig of 683 bp in length
* 31112 31211: gap of 100 bp
* 31212 31931: contig of 720 bp in length
* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length

* 15913 16012: gap of 100 bp
* 16013 16731: contig of 719 bp in length
* 16732 16831: gap of 100 bp
* 16832 17506: contig of 675 bp in length
* 17507 17606: gap of 100 bp
* 17607 18339: contig of 733 bp in length
* 18340 18439: gap of 100 bp
* 18440 19109: contig of 670 bp in length
* 19110 19209: gap of 100 bp
* 19210 19900: contig of 691 bp in length
* 19901 20000: gap of 100 bp
* 20001 20693: contig of 693 bp in length
* 20694 20793: gap of 100 bp
* 20794 21464: contig of 671 bp in length
* 21465 21564: gap of 100 bp
* 21565 22245: contig of 681 bp in length
* 22246 22345: gap of 100 bp
* 22346 23049: contig of 704 bp in length
* 23050 23149: gap of 100 bp
* 23150 23857: contig of 708 bp in length
* 23858 23957: gap of 100 bp
* 23958 24685: contig of 728 bp in length
* 24686 24785: gap of 100 bp
* 24786 25499: contig of 714 bp in length
* 25500 25599: gap of 100 bp
* 25600 26302: contig of 703 bp in length
* 26303 26402: gap of 100 bp
* 26403 27108: contig of 706 bp in length
* 27109 27208: gap of 100 bp
* 27209 27910: contig of 702 bp in length
* 27911 28010: gap of 100 bp
* 28011 28720: contig of 710 bp in length
* 28721 28820: gap of 100 bp
* 28821 29484: contig of 664 bp in length
* 29485 29584: gap of 100 bp
* 29585 30328: contig of 744 bp in length
* 30329 30428: gap of 100 bp
* 30429 31111: contig of 683 bp in length
* 31112 31211: gap of 100 bp
* 31212 31931: contig of 720 bp in length
* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length
* 39913 40012: gap of 100 bp
* 40013 40712: contig of 700 bp in length
* 40713 40812: gap of 100 bp
* 40813 41514: contig of 702 bp in length
* 41515 41614: gap of 100 bp
* 41615 42294: contig of 680 bp in length
* 42295 42394: gap of 100 bp
* 42395 43094: contig of 700 bp in length
* 43095 43194: gap of 100 bp
* 43195 43915: contig of 721 bp in length
* 43916 44015: gap of 100 bp
* 44016 44740: contig of 725 bp in length
* 44741 44840: gap of 100 bp

* 44841 45533: contig of 693 bp in length
* 45534 45633: gap of 100 bp
* 45634 46330: contig of 697 bp in length
* 46331 46430: gap of 100 bp
* 46431 47137: contig of 707 bp in length
* 47138 47237: gap of 100 bp
* 47238 47927: contig of 690 bp in length
* 47928 48027: gap of 100 bp
* 48028 48709: contig of 682 bp in length
* 48710 48809: gap of 100 bp
* 48810 49509: contig of 700 bp in length
* 49510 49609: gap of 100 bp
* 49610 50331: contig of 722 bp in length
* 50332 50431: gap of 100 bp
* 50432 51135: contig of 704 bp in length
* 51136 51235: gap of 100 bp
* 51236 51981: contig of 746 bp in length
* 51982 52081: gap of 100 bp
* 52082 52783: contig of 702 bp in length
* 52784 52883: gap of 100 bp
* 52884 53580: contig of 697 bp in length
* 53581 53680: gap of 100 bp
* 53681 54398: contig of 718 bp in length.

FEATURES

Location/Qualifiers
1..54398
/organism="Homo sapiens"

Query Match 5.9%; Score 311; DB 73; Length 54398;
Best Local Similarity 100.0%; Pred. No. 2.6e-156;

Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3819 agagccatcgctcagaaatctgatgaacagcagtgccgcaaaaagggtcaaaagagctcaaa 3878

Db 29407 AGAGGCCATACGGCTTCAGAAATCTGTGAACAGCAGTGGCTGTGAGGAAAGAGCTCAA 29348

QY 3879 gaagatatattagaaatgaagatgaacagaatagtcgcgcaaaaagggtcaaaagagc 3938

Db 29347 GAAGATATATTAGAAATGAAGATGAACAGANTAGTCCGCCAAAAGGGTAAAGAGGC 29288

QY 3939 cgaccaccaaaaccttggtgaggtacaccaaaagaagccaaatgaaaactct 3998

Db 29287 CGACCACCAAAAGCTCTTGTGTGGAGGTACACCAAAAAGAGCCAAACATGAAACTTCT 29228

QY 3999 aaaaaaggaagcaaaaaaatactgacctccagcaccagagaggaagaagaagaa 4058

Db 29227 AAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCAGAGGAGGAGGAAGAAGAA 29168

QY 4059 agacaaagtgaatacggacagagaagtcacaaagcaaacagaccagtgctcaagaga 4118

Db 29167 AGACAAAGTGAATAACGGAACAGAGAGTCCAAAAGCAACAGCACCAGTGCTCAAGGAGA 29108

QY 4119 gcacagcagag 4129

Db 29107 GCACAGCAGAG 29097

RESULT 10

AC068224

LOCUS

DEFINITION

Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS

SEQUENCE SAMPLING.

AC068224

VERSION

AC068224.1 GI:7671284

KEYWORDS

HTG; HTGS_PHASE0.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 54398)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homosapiens chromosome 3, clone RP11-660H19

Unpublished

REFERENCE

2 (bases 1 to 54398)

* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 757: contig of 757 bp in length
758 857: gap of 100 bp
858 1635: contig of 778 bp in length
1636 1735: gap of 100 bp
1736 2490: contig of 755 bp in length
2491 2590: gap of 100 bp
2591 3343: contig of 753 bp in length
3344 3443: gap of 100 bp
3444 4224: contig of 781 bp in length
4225 4324: gap of 100 bp
4325 5088: contig of 764 bp in length
5089 5188: gap of 100 bp
5189 5968: contig of 780 bp in length
5969 6068: gap of 100 bp
6069 6844: contig of 776 bp in length
6845 6944: gap of 100 bp
6945 7701: contig of 757 bp in length
7702 7801: gap of 100 bp
7802 8573: contig of 772 bp in length
8574 8673: gap of 100 bp
8674 9430: contig of 757 bp in length
9431 9530: gap of 100 bp
9531 10295: contig of 765 bp in length
10296 10395: gap of 100 bp
10396 11139: contig of 744 bp in length
11140 11239: gap of 100 bp
11240 11988: contig of 749 bp in length
11989 12088: gap of 100 bp
12089 12874: contig of 786 bp in length
12875 12974: gap of 100 bp
12976 13750: contig of 776 bp in length
13751 13850: gap of 100 bp
13851 14638: contig of 788 bp in length
14639 14738: gap of 100 bp
14739 15520: contig of 782 bp in length
15521 15620: gap of 100 bp
15621 16381: contig of 761 bp in length
16382 16481: gap of 100 bp
16482 17248: contig of 767 bp in length
17249 17348: gap of 100 bp
17349 18123: contig of 775 bp in length
18124 18223: gap of 100 bp
18224 18977: contig of 754 bp in length
18978 19077: gap of 100 bp
19078 19860: contig of 783 bp in length
19861 19960: gap of 100 bp
19961 20732: contig of 772 bp in length
20733 20832: gap of 100 bp
20833 21590: contig of 758 bp in length
21591 21690: gap of 100 bp
21691 22446: contig of 756 bp in length
22447 22546: gap of 100 bp
22547 23333: contig of 787 bp in length
23334 23433: gap of 100 bp
23434 24212: contig of 779 bp in length
24213 24312: gap of 100 bp
24313 25093: contig of 781 bp in length
25094 25193: gap of 100 bp
25194 25958: contig of 765 bp in length
25959 26058: gap of 100 bp
26059 26839: contig of 781 bp in length
26840 26939: gap of 100 bp
26940 27705: contig of 766 bp in length
27706 27805: gap of 100 bp
27806 28568: contig of 763 bp in length
28569 28668: gap of 100 bp
28669 29419: contig of 751 bp in length

29420 29519: gap of 100 bp
29520 30311: contig of 792 bp in length
30312 30411: gap of 100 bp
30412 31202: contig of 791 bp in length
31203 31302: gap of 100 bp
31303 32082: contig of 780 bp in length
32083 32182: gap of 100 bp
32183 32934: contig of 752 bp in length
32935 33034: gap of 100 bp
33035 33796: contig of 762 bp in length
33797 33896: gap of 100 bp
33897 34683: contig of 787 bp in length
34684 34783: gap of 100 bp
34784 35552: contig of 769 bp in length
35553 35652: gap of 100 bp
35653 36406: contig of 754 bp in length
36407 36506: gap of 100 bp
36507 37294: contig of 788 bp in length
37295 37394: gap of 100 bp
37395 38159: contig of 765 bp in length
38160 38259: gap of 100 bp
38260 39017: contig of 758 bp in length
39018 39117: gap of 100 bp
39118 39890: contig of 773 bp in length
39891 39990: gap of 100 bp
39991 40774: contig of 784 bp in length
40775 40874: gap of 100 bp
40875 41643: contig of 769 bp in length
41644 41743: gap of 100 bp
41744 42521: contig of 778 bp in length
42522 42621: gap of 100 bp
42622 43393: contig of 772 bp in length
43394 43493: gap of 100 bp
43494 44268: contig of 775 bp in length
44269 44368: gap of 100 bp
44369 45149: contig of 781 bp in length
45150 45249: gap of 100 bp
45250 46024: contig of 775 bp in length
46025 46124: gap of 100 bp
46125 46868: contig of 744 bp in length
46869 46968: gap of 100 bp
46969 47720: contig of 752 bp in length
47721 47820: gap of 100 bp
47821 48596: contig of 776 bp in length
48597 48696: gap of 100 bp
48697 49462: contig of 766 bp in length
49463 49562: gap of 100 bp
49563 50335: contig of 773 bp in length
50336 50435: gap of 100 bp
50436 51207: contig of 772 bp in length
51208 51307: gap of 100 bp
51308 52078: contig of 771 bp in length
52079 52178: gap of 100 bp
52179 52954: contig of 776 bp in length
52955 53054: gap of 100 bp
53055 53813: contig of 759 bp in length
53814 53913: gap of 100 bp
53914 54675: contig of 762 bp in length
54676 54775: gap of 100 bp
54776 55537: contig of 762 bp in length
55538 55637: gap of 100 bp
55639 56413: contig of 776 bp in length
56414 56513: gap of 100 bp
56514 57278: contig of 765 bp in length
57279 57378: gap of 100 bp
57379 58149: contig of 771 bp in length
58150 58249: gap of 100 bp
58250 59023: contig of 774 bp in length
59024 59123: gap of 100 bp
59124 59906: contig of 783 bp in length
59907 60006: gap of 100 bp
60007 60792: contig of 786 bp in length
60793 60892: gap of 100 bp


```

Db 26351 GATCAATAAAATCTGTAATTTGAATGAGT 26381
|||||
RESULT 7
AK026889 2079 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens CDNA: FLJ23236 fls, clone COL00725.
ACCESSION AK026889
VERSION AK026889.1 GI:10439854
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE NEDO human cDNA sequencing project
ORGANISM Homo sapiens
REFERENCE
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2079)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
DIRECT SUBMISSION
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leins.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source Location/Qualifiers
1. 2079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL00725"
/clone_lib="COL"
/tissue_type="colon"
/notes="cloning vector pME18SFL3"
304..681
/notes="unnamed protein product"
/codon_start=1
/db_xref="BAB15584.1"
/translation="MDLITKLIVQKPKGSQRKRGHTASEDEQOWPPEKRLKEDI
LENEDEQNSPPKGRGRPPKPLGGTPEKEPTMTKTSKSKSKSKSPAPPEEEEEE
QSGNTEQKSKSHRVSRRAQR"
BASE COUNT 759 a 354 c 339 g 627 t
ORIGIN

Query Match 8.4%; Score 441; DB 89; Length 2079;
Best Local Similarity 100.0%; Pred. No. 9.Be-227;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3689 gctcgaatggagagcctagaggcaggaagaaaccccgctcacagacaggaggagaa 3748
|||||
Db 237 GTCCTGATTTGGAGAGCCTAGAGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 296
|||||
QY 3749 attagtgatgaatgaactgaactgaactgaactgaactgaactgaactgaactgaact 3808
|||||
Db 297 ATTAGTGATGGATGACTTGACTTAAGTTGGTACAGGAACAGAAACCTTAAGGCAGTCAGCG 356
|||||
QY 3809 aagtcggaagagccatagcgttcagatctctgatgaacagcagtgccctgaggagaaa 3868
|||||
Db 357 AAGTCGGAAGAGAGCCATACGCGCTTCAGATCTCATGAACAGCAGTCGCCCTGAGGAGAA 416
|||||

```

```

QY 3869 gaggtcacaagaagatatattagaaatgaagatgaacagaatagtcgcccaaaaaaggg 3928
|||||
Db 417 GAGGCTCAAAAGAGATATATTAGAAATGAAGATGAACAGAATAGTCCGCCCAAAAAGGG 476
|||||
QY 3929 taaaagagccgacacacacacacacacacacacacacacacacacacacacacacacacac 3988
|||||
Db 477 TAAAAGAGGCCACACACACACACACACACACACACACACACACACACACACACACACAT 536
|||||
QY 3989 gaaaacttctaaaaaagcagcaaaaaaaatctggacctccagcaccagagagagagga 4048
|||||
Db 537 GAAACTTCTAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 596
|||||
QY 4049 agaagaagaagacacacacacacacacacacacacacacacacacacacacacacacacac 4108
|||||
Db 597 AGAAGAAGAAGACAAAGTGGAAATACGGAACAGAAAGTCCAAAAGCAACACAGCAGCGAGT 656
|||||
QY 4109 gtcaagagagagacacacacacacacacacacacacacacacacacacacacacacacac 4129
|||||
Db 657 GTCAAGGAGAGCAGCAGCAGAG 677
|||||
RESULT 8
AC016449/c
LOCUS AC016449 72157 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016449
VERSION AC016449.2 GI:9104517
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 72157)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-45L14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72157)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,A., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6479175.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1189
Center clone name: 45_L14
-----
* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for

```



```
QY 1849 atcctttctggaatgatcaaaatcttcttcttgagagtagcacctgtgcacatagata 1908
DB 721 ATCTTTCTCTGGAATGATCAAGTTCTTCTTGAGAGGATAGCACCTGTGCACATAGATA 780
QY 1909 cgaatctatcagtgcttcttattaaacaagtgaaacaaataatagatggaacagcagtg 1968
DB 781 CGAAATCTATCAGTGTCTTATTAAACAAGTGAAACAATCAATAGATGGAACAGCAGATG 840
QY 1969 atgaagatgaggtgttccaatgatcaagccatcagcagcagcagcagcagcagcagc 2028
DB 841 ATGAAGATGAGGGTGTTCCACATGATCAAGCCATCAGAGAGGCTTGAACCTGCTTAAG 900
QY 2029 tactctcattacacatcccatctcattcttctctgctgaacatttgaaatcattctgg 2088
DB 901 TACTCTCATTTACACATCCATCTCATTTTCATTTCTGCTGAAACATTTGAATCATTTACTGG 960
QY 2089 ctgtctgaaatgtagatgataaaagtgagcagaagcagcagcagcagcagcagcagcagc 2148
DB 961 CTTGTCTGAAATGGATGATGAAAAAGTAGCAGAAAGTAGCAGAAAGTAGCAGAAATTTTCAAAAACA 1020
QY 2149 caggaagcaaaattgaagaggtttttccacacacacacacacacacacacacacacac 2208
DB 1021 CAGGAAGCAAAATTTGAAGAGGATTTTCCACATCATCATGATCAGCTTGTCTGCTGTTTTAC 1080
QY 2209 atcacaatataaaaggaagcccccgcg-tcaagccaaatatgccattcattgtatccat 2267
DB 1081 ATCACAATCTTAAAAAGGACCCCGCTCAAGCCAAATATGCCATTCATTCGTATCCAT 1140
QY 2268 gcgatatttctagttaagagaccagcttggcagagatatattgagccctgcataagagc 2327
DB 1141 GCGATATTTTCTAGTAAGAGACYCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGC 1200
QY 2328 ctatgaccagcaacctgaaacatctcataaacaccattggttactattggtcatttgc 2387
DB 1201 CTAGATCCCAAGCAACCTGGAACATCTCATTAACACCATTGTTGTTACTATTTGTCATATTGCT 1260
QY 2388 ctctgtcacctgatcaatttgccttcttcttcttcttcttcttcttcttcttcttcttct 2447
DB 1261 CTCCTTGCACCTGATCAATTTGCTGCTCTCTTGAATCTTTGTTAGTACTTCTTCAATTTG 1320
QY 2448 aaagatcttctatgatgatcagcttcttcttcttcttcttcttcttcttcttcttcttct 2507
DB 1321 AAAGATCTTCTATGATGATGCTGCTTCCAGGAAAAAGACAACTAAACTTTGGGTTCCTCA 1380
QY 2508 gatgaagaagtatctctgagacaatggcacaattcagcagctattaaatgatggttcga 2567
DB 1381 GATGAGAAGTATCTCTGAGACAAATGGTCAAAATTCAGCTATTAAATGATGTTGTCGA 1440
QY 2568 tggctacttggg 2578
DB 1441 TGCTACTTGG 1451
```

```
RESULT 5
LOCUS HS49J10 137246 bp DNA PRI 22-NOV-1999
DEFINITION Human DNA sequence from PAC 49J10, BRCA2 gene region chromosome
13q12-13 contains ESTs.
ACCESSION 284572
VERSION 284572.1 GI:1813974
KEYWORDS 13q12-13.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137246)
Whiteley, M.
Direct Submission
Submitted (09-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
IMPORTANT:
This sequence is not the entire insert of clone . It may be
```

shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 49J10 is at 1 in this sequence. The true right end of clone 267P19 is at 43908.

The true left end of clone 179J15 is at 137143.

49J10 is from the human PAC library described in Ioannou A.P. et al Nature Genet 6, 84-89.

```
FEATURES
source
    1..137246
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="13"
        /map="13q12-13"
        /clone="XX-49J10"
        /note="AluX repeat: matches 281..1 of consensus;
        incomplete repeat"
repeat_region
    1551..1831
        /note="17 copies of 2 mer 100 % conserved"
repeat_region
    1838..1871
        /note="LIPAI5 repeat: matches 893..1 of consensus"
repeat_region
    1882..2741
        /note="LIPAI6 repeat: matches 5389..5139 of consensus"
repeat_region
    2597..2848
        /note="MER44A repeat: matches 247..1 of consensus"
repeat_region
    3149..3405
        /note="AluSc repeat: matches 188..120 of consensus;
        incomplete repeat"
repeat_region
    3853..3922
        /note="MIR repeat: matches 58..262 of consensus"
repeat_region
    5620..5810
        /note="AluX repeat: matches 1..301 of consensus"
repeat_region
    7829..8133
        /note="AluX repeat: matches 291..1 of consensus"
repeat_region
    10957..11242
        /note="LIPAI6 repeat: matches 903..842 of consensus"
repeat_region
    12344..12405
        /note="LIPAI6 repeat: matches 833..68 of consensus"
repeat_region
    12381..13119
        /note="HY1 repeat: matches 110..1 of consensus"
repeat_region
    1428..14540
        /note="MIR2 repeat: matches 126..1 of consensus"
repeat_region
    14951..15071
        /note="AluX repeat: matches 302..4 of consensus"
repeat_region
    16589..16885
        /note="LIPAI6 repeat: matches 3807..4127 of consensus"
repeat_region
    18067..18397
        /note="MLT1A2 repeat: matches 374..1 of consensus"
repeat_region
    19076..19481
        /note="MLT1A2 repeat: matches 374..1 of consensus"
repeat_region
    19516..19799
        /note="AluX repeat: matches 19..302 of consensus;
        incomplete repeat"
repeat_region
    19808..20127
        /note="AluX repeat: matches 2..299 of consensus"
prim_transcript
    <21359..26899
        /note="match: multiple ESTs; match: H4781 D20082 W20070
        T64867 H45693; match: F12354 N91467 N83765 H45741 H93424"
repeat_region
    24795..25089
        /note="AluX repeat: matches 294..1 of consensus"
repeat_region
    26083..26106
        /note="12 copies of 2 mer 96 % conserved"
prim_transcript
    complement(27120..27476)
        /note="match: multiple ESTs; match F09972 T72607"
prim_transcript
    <27633..28101
        /note="match: multiple ESTs; match: 5' EST T51695 clone
        72405; Paired with EST T51543 matching this clone; match
        H91918"
```



```

|||||
Db 3361 GAATGAAACTGTAAAGCAATGAAGCAGCAGCTCAATCAAGCTCTCTCGAAGAATAA 3420
QY agggagagcttgaatctgaatgaatgatacagatgaataaagattacacaatctt 3636
Db agggagagcttgaatctgaatgaatgatacagatgaataaagattacacaatctt 3480
QY cacccttgcggggaataaagtgacaagagagcactctgactctgtaaggtctgaat 3696
Db CACCTTTGCGGGGAAAAAAGTGACAAGAGAGAGCAGCTGATCTGTGAAGTCTGAAT 3540
QY tggagaagctagaggcaggaataaaccgcccgctacagaaacagggggaataaggta 3756
Db TGGAGAAGCCTAGAGCAGGAAAAAACCCTGCTCAGAACACAGGAGGAGAAATAGGTA 3600
QY tggatgactgactaagttgggtacaggaacagaaacctaagagcagtcagcgaagtcga 3816
Db TGGATGACTTGACTAAGTTGGTACAGAGACAGAAACCTAAGAGCAGTCAGCGAAGTCGGA 3660
QY aaagagccatcagcggcttcagaaatctgataacagcagtgagcctgaggaagaggtca 3876
Db AAAGAGCCATACCGCTTCAGAAATCTGATCAACAGCAGTGGCCTGAGGAAAGAGGCTCA 3720
QY aagaagatatattagaataatgaatgaacagaaatagtcgcgcaaaaagggtaaaagag 3936
Db AAGAAGATATATTAGAATAATGAAGATGAACAGAAATAGTCCGCCAAAAAAGGGTAAAGAG 3780
QY ggcagaccacaaaacctctgtgagggtacacaaaagaagagccaaacaaatgaaacatt 3996
Db GCGGACCACCAAAACCTCTGTGTGGAGGTACACCAAAAGAGAGCCAAACATGAAAACTT 3840
QY ctaaaaaggaagcaaaaaatctggacctccagcaccagagggagggaagaagaag 4056
Db CTAAGAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAG 3900
QY aaagcaaaagtggaaatacggaaacagaaagtcacaaagcaaaagcagccagtgtaagga 4116
Db AAAGCAAAAGTGGAAATACGGACACAGAAATCCAAAAGCAACAGCACCAGGTGTCAAGGA 3960
QY gagcagcagagagcagaaatctctgaatctagtgcaatgaatccacacagtcacacac 4176
Db GAGCAGCAGAGAGCAGAAATCTCTGATCTAGTCAATGTAATCCACAGTCACACAC 4020
QY cacagaaagcagaggaagaccatcaaaacgcccacacacacacacacacacacacacac 4236
Db CACAGAAAGGACGAGGAAGACCATCAAAACGCGCATCACCATCACAACCAAAAAAATG 4080
QY 4237 t 4237
Db 4081 t 4081

```

```

RESULT 4
LOCUS HSU050533 1852 bp mRNA PRI 27-NOV-1996
DEFINITION Human BRCA2 region, mRNA sequence CG008.
ACCESSION U050533
VERSION U050533.1 GI:1531605
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1852)
AUTHORS Couch,F.J., Rommens,J.M., Neuhausen,S.L., Belanger,C., Dumont,M.,
Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L.,
Farid,L., Frye,C., Hattier,T., Janecki,T., Jiang,P., Kehrer,R.,
Leblanc,J.-F., McArthur-Morrison,J., McSweeney,D., Miki,Y.,
Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M.,
Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T.,
Tranchant,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H.,
Labrie,F., Skolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L.,
Tavtigian,S.V. and Simard,J.

```

```

TITLE Generation of an integrated transcription map of the BRCA2 region
on chromosome 13q12-q13
JOURNAL Genomics 36 (1), 86-99 (1996)
MEDLINE 96411650
REFERENCE 2 (bases 1 to 1852)
AUTHORS Simard,J.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular
Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,
Quebec City, Quebec G1V 4G2, Canada
FEATURES
source
1..1852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
/notes="CG008; DSEG numbers: D13S171 and D13S310"
BASE COUNT 609 a 348 c 379 g 505 t 11 others
ORIGIN

```

```

Query Match 21.4%; Score 1126; DB 97; Length 1852;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1446; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 1129 ttaaagtgaggtcacatgacccctgaggaagctattctctggtcaatgatacacttaattgtgtagag 1188
Db 1 TTAAAGTGAAGTCAATGATGACCTTGAGGAGCTATTAGACATGATGTTATTTGTCTCAATAG 60
QY 1189 ttacagctctcaaaaaggattattctctggtcaatgatacacttaattgtgtagag 1248
Db 61 TTACAGCTCTCAAAAAGGATATTCTTCTGTGTCATGATGATGTTATTTGTGAGAG 120
QY 1249 agagaacattagacaacagatgagagtagcgaagaagcagcagtgaggactggccaaa 1308
Db 121 AGAGAACATTAGACAAACCATGGAGAGTAGGCAAAAGCCATGATGGGACTTGGCCAAA 180
QY 1309 ttataagaataatgctttcacagtcagcagctggaagaagatgctgcaaacagatagcat 1368
Db 181 TTTATTAAGAAATATGCTTTTACAGTCAGCAGCTGGAAAGATGCTGCAAAACAGATAGCAT 240
QY 1369 ggatacaaaagcaaaattgctacatatataattatacaaaatagattgatgactgactctg 1428
Db 241 GGATCAAGACAAAATTTGCTACATATATATATCAAAAATAGTATTGATGATCGACTACTTG 300
QY 1429 ttgaacggtcttctgctcaatcagatgctcctcacaatttagaaactcagacagcagta 1488
Db 301 TTGAACGGATCTTGTCTCAATACATGCTTCCCTCACAAATTTAGAAACTACAGACGGATGA 360
QY 1489 aatgcttattacttgatgcccacactggattaaatgctgtgaagcattgaaatgaa 1548
Db 361 AATGCTTATATTACTTGTATGTCACACTGGATTTAAATGCTGTGAAGCATTTGAATGAA 420
QY 1549 tgtgaaaatgcataaactgctccgacatcaagtaaggattgctgactgacttaagc 1608
Db 421 TGTGGAATGTCAAAATTCGCTCCGACATCAAGTAAGGATTTGCTTGACTTGATTAAAG 480
QY 1609 aacccaaacagatgacagtgctcaagggccatattttcaaaagtgtggttattacaagaa 1668
Db 481 AACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTACAAGAA 540
QY 1669 attacctgatctctgtaaggctcaggatttcattgaagaataatcacacaggtgttagaag 1728
Db 541 ATTACTGATCTCTGTTAAAGGCTTCAGGATTTTCATGAAGAAATTCACACAGGCTGTGAAG 600
QY 1729 atgatagaataaagaacagattagaagctacttctttagtccaacactgctcctgaagc 1788
Db 601 ATGATCAGAAAATAAAGAAAGCAGTGTAGAGTACTTGTAGTCAACATGCTCTCTGCAAGC 660
QY 1789 aggctgaaggttctgctgctgaaataactaaagaagttgggcaaccccaaacacagcctacaa 1848
Db 661 AGGCTGAAGGTTGTGCGTGAATTAACTAAGAAGTTGGCAACCCCAACACACGCTTACAA 720

```


Db 2282 AAAAGGACCCCGTCAAGCCAAATATGCAATTCATTTGTATCCATCGCATATTTTCTA 2341
Qy 2281 gtaagagaccagttgacagatattgagctctgcataaagagcctagatccaagca 2340
Db 2342 GTAAGGACCCAGTTGACAGATATTGAGCCTCTGCATAAAGAGCCTAGATCCAAGCA 2401
Qy 2341 acctggaacatctcataacaccattggttactattggtcattggtcctctctgacactg 2400
Db 2402 ACCTGGAACATCTCATACACCATTTGTTACTATTGGTCATATTGGCTCTCTTGCACCTG 2461
Qy 2401 atcaattgtctctctggaaactctgggttagctacttccattgtgaaagattcttca 2460
Db 2462 ATCAATTTGCTGCTCTTGAATCTTTGGTAGTACTTTTCATTTGTAAGATCTTCTCA 2521
Qy 2461 tgaatgatcggtctccagggaagaaagcaactaaactttgggttccagatgaagaatat 2520
Db 2522 TGAATGATCGGCTTCCAGGGAAGAAAGACACTAAACTTTTGGGTTCCAGATGAAGAATAT 2581
Qy 2521 ctcttgagacaattggtcaaaattcaggctatttaaaatgtaggttcgattggttcttgaa 2580
Db 2582 CTCTTGAGACAAATGCTCAAAATTCAGGCTATTAAATGATGGTTTCGATGGCTACTTGGAA 2641
Qy 2581 tgaataataatcacagtaaatcaggaactctaccttaagattgctaaacaataatgc 2640
Db 2642 TGAATAATAATCACAGTAATAATCAGGAACCTTCTACCTTAAGATTGCTAACACAATAATGC 2701
Qy 2641 atagtgtgagacttgacagaaacaggggaaaaattagtaaaacagatatgtcacgcttga 2700
Db 2702 ATAGTGATGGAGACTTGACAGAAACAGGGGAAATTTAGTAAACAGATGATGTCAGCTGTA 2761
Qy 2701 gacttgcctggtgagtgctattgagctggcacagaaacctgtttaccatgaatca 2760
Db 2762 GACTTGTCTGGAGTGCTATTGTGAAGCTGGCACAAAGACCTGTGTACCTGAATCA 2821
Qy 2761 tcacattgaacaataacagctatgtgaattagctatcaacagatgaatgtctatcaagtaa 2820
Db 2822 TCACATTAGAACAATATCAGCTATGTGCTATGCTATCAACGATGAATGCTATCAAGTAA 2881
Qy 2821 gacaagtgttggccagaaacttcacaaagcctttcccggtttagcgcttccacttgagt 2880
Db 2882 GACAAGTGTTTGGCCAGAACTTCACAAAGGCCCTTCCCGTTTACGGCTTCCACTTGAGT 2941
Qy 2881 atatggcaactgtgccccttctgtcaaaagatcctgttaaaagagagagagctcatgcta 2940
Db 2942 ATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAGGAGAGAGAGCTCATGCTA 3001
Qy 2941 ggaatgttgggtgaaaaataaataatgaagcgggagatctgaagcagcatgcagctg 3000
Db 3002 GGCAATGTTTGGTGAAAAATATAATGTAAGCGGGAGTATCTGAAGCAGCATGCAGCTG 3061
Qy 3001 ttagtgaataatttctctctacagagtagtattgtttccatatatacaattcaccttt 3060
Db 3062 TTAGTGAATAATTTGTCTCTTCTACAGAGTATGTTGTTCCATATACAAATTCACCTTT 3121
Qy 3061 tggcacatgacccagattatgtcaaatgacaggatattgaacaacttaaaagttaaaag 3120
Db 3122 TGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACCTTAAAGATTTAAAG 3181
Qy 3121 aatgtcttgggttcttcgaaatattaatggtcaaaaatgaaataaacaagtcacgctt 3180
Db 3182 AATGCTTTGGTTTGTCTGAAATATTAAATGGCTTAAAAATGAAATAAACAGTCACGCTT 3241
Qy 3181 ttatcaagaagatgtgagaataattaaacaaacaaagatgccaaagaccagatgatg 3240
Db 3242 TTATCAGAAAGATGTTAGAAATATTAAACAAACAAAGATGCCAAGGACGATGATG 3301
Qy 3241 caaaaaatgaataaagaaactgtacactgtgtgtgattggtccatgaataatcatcatgtaa 3300
Db 3302 CAAAAATGAATGAATAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAA 3361
Qy 3301 agagtactacacagtttgaaactctcctaaagaccgggtactaccagctcgtttcttca 3360
Db 3362 AGAGTACTACATAGTTTGGAAATCTCCTAAAGACCCGGTACTACCAAGCTCGTTCTTCA 3421

Qy 3361 ctcaacctgacaagaatttcagtaacaccacaaattatctcctcctgaaatgaatcat 3420
Db 3422 CTCAACCTGACAGAATTTTCAGTAAACACCAAAATTTATCTGCCCTCTGAAATGAATCAT 3481
Qy 3421 ttttcaacctctggaaaacccctaaacaaatgcttagggagctgttaacaagccacttt 3480
Db 3482 TTTTCACTCTCTGGAAAACCTTAAACAAACCAATGTTCTAGGAGCTGTTTAAACAGCCACTT 3541
Qy 3481 catcagcaggcaagaatctctcagaccacaaatcatcacgaatggaactgtgaagaatgcaa 3540
Db 3542 CATCAGCAGGCAACAATCTCAGACCAATCATCAGCAATGCAAACTGTAAGCAATGCAA 3601
Qy 3541 gcacagactcaaatccaaactctcctgagaataaaagggagctgtgattgattctctgaa 3600
Db 3602 GCACAGCTCAATCCAAAGCTCTCTGGAAGAAATAAGGGAGGCTGTGATAGTCTTGAAA 3661
Qy 3601 tggatcagtgaaaaatgaagattacacaaatgtcttcacctttccgggggaaaaaagt 3660
Db 3662 TGGATCAGTGAAAATGAAGATTACACAATGTCTTCACTTTGCGGGGAAAAAAGTGT 3721
Qy 3661 acaagagagacgactctgactcttgaaggtctgaattggaagccttagagcgaggaataa 3720
Db 3722 ACAAGAGAGAGCAGCTCTGATCTTGAAGTCTGAATTTGGAGAAGCCTTAGAGCGCAGGAAA 3781
Qy 3721 aaacgcccgtcacagacagagagagaaattagattgattgactgaactgaattggttac 3780
Db 3782 AAACGCCGTGACAGACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTAC 3841
Qy 3781 aggaacgaacacctaaaggcagtcagcgaactcgaaaagggccatacgccttcagaat 3840
Db 3842 AGGAACGAACCTTAAAGGCAGTCAGCGAAGTCGGAAGAGGCGCATACGCGCTTCAGAAT 3901
Qy 3841 ctgatatgaacagctgacctgaggaagaagggctcacaagaagatatctagaagaatgaag 3900
Db 3902 CTGATGAACAGCAGTGGCTGAGGAAAAGAGGCTCAAGAAGAGATATATTAGAAAATGAAG 3961
Qy 3901 atgaacgaatagtcgcccaaaaagggttaaaagagccgaccacaaacctctgtgtg 3960
Db 3962 ATGAACAGATAGTCCGCCAAAAAGGTTAAAGAGCCGACACCAAAACCTCTTGTGTG 4021
Qy 3961 gagtacaccaaaaagaagagccaaactgaaaacttctaaaaaaggaagcaaaaaaaat 4020
Db 4022 GAGGTACACCAAAAAGAGAGGCCAACAAATGAAAACCTTCTAAAAAGGAAGCAAAAAAAT 4081
Qy 4021 ctggaacctcaagcaccagagagaggaagaagaagaagaacaaagtggaaatacaggaac 4080
Db 4082 CTGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAAGACAAAGTGGAAATACGGAAC 4141
Qy 4081 agaagtcacaaagcaaacagcagcagctgtcaagagagcagcagcagagcagaaatctc 4140
Db 4142 AGAAGTCCAAAGCAACAGCAGCCAGTGTCAAGGAGAGCAGCAGAGAGCAGAGATCTC 4201
Qy 4141 ctgaatctagtgtcaattgaaatccacagctccacaccagaaagagcaggaagaccat 4200
Db 4202 CTGAATCTAGTGAATTTGAATCCACAGTCCACACAGTCCACAGAAAGGACGAGGAAGACCAT 4261
Qy 4201 caaaaaagccatcacccatcaacacaaaaaaatgt 4237
Db 4262 CAAAAACCCCATCCCATCACACCAACCAAAAAAATGT 4298

RESULT 3

AB023196
LOCUS AB023196 5110 bp mRNA PRI 16-JUN-1999
DEFINITION Homo sapiens mRNA for KIAA0979 protein, partial cds.
ACCESSION AB023196
VERSION AB023196.1 GI:4589601
KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBbluescriptII
SOURCE SK plus clone:hj07056.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

QY 61 ctgtcatggtctatctaaagactagagaccaatgatggaataatcacatatcccgctgggg 120
DB 122 CTGTCTATGGCTCAATTCAAGACTAGGACCAATGATGGAATAATTACATATCCGCTGGG 181
QY 121 tcaagaaatcatcagataaataatctaaagagagagatggtgagacattaaagatggtg 180
DB 182 TCAAGGAATATCAGATAAATAATATCTAAAGAGGAGATGTTGACGAGATTAAGATGGTTG 241
QY 181 tgaactttttatggatggaccagactctgaagaagaagaaggagctttatttaaacc 240
DB 242 TGAATAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTAAACC 301
QY 241 tagctttacatctgtctcaagattttttctcaagcatccctgtaagatggttcogcttac 300
DB 302 TAGCTTTACATCTGCTTCAGATTTTTTCTCAAGCATCCTGATAAGATGTTTCGCTTAC 361
QY 301 tggtagcctgctcctgctgatattttcaaggatttatgctcctgaagctcccttacacat 360
DB 362 TGGTAGCCTGCTGCTGCTGATATTTTCAGGATTTATGCTCTGGAAGCTCCTTACACAT 421
QY 361 ccctgtataaactaaaggatatatttattgtttatacaagacaggttgaagggtctagagg 420
DB 422 CCCTGATAAACTAAAGGATATATTATGTTTATAACAAGACAGTTGAAGGGCTAGAGG 481
QY 421 atacaagagccacaaatccaataggtattttttatttacttgagaacattgcttgggtca 480
DB 482 ATACAAGAGCCACAAATTCATAGGTATTTTTTATTACTTGAGAACAATTTGCTGGGTCA 541
QY 481 agtcataacatatgtcttgagttgagatagcaatgaattttccaccagctataca 540
DB 542 AGTCATATACATATGCTTTGAGTTTAGAGATAGCAATGAAATTTTCACCCAGCTATACA 601
QY 541 gaaccttatttccagtataaacaatggccacaaatcagaagaatcccatatgacatggtag 600
DB 602 GAACCTTATTTTCAGTTATAAACAATGGCCACAATCAGAAGTCCATATGCACATGGTAG 661
QY 601 acccttagctctattatttgaagggtgatacagtgctcagggagcttttgatacgg 660
DB 662 ACCTTAGAGCTCTATTTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGG 721
QY 661 ttttagtaaatctggtacctgctcataagaatttaacaagcaagcatatgatttggcaa 720
DB 722 TTTTAGTAATCTGGTACCTGCTCATAGNAATTTAAACAAGCAAGCATATGATTGGCAA 781
QY 721 aggtcttaactgaagagagcagctcaagctattgagccatatattaccacttttttaatc 780
DB 782 AGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTAATC 841
QY 781 aggtctctgattggtggaataatcatcagcatttgcagagcatgctttgacttaa 840
DB 842 AGGTTCTGATGCTGGGAACAATCATCAGCGATTTGTGAGAGCATGCTTTTGACTTAA 901
QY 841 ttttgagctctacaataattgaagtcatttgcgtctctctgttttaccacagcttgaat 900
DB 902 TTTTGGAGCTCTACATATTGATAGTCAATTGCTGCTCTGTTTTACCCACGCTTGAAT 961
QY 901 ttaaaataaagagcaatgataatgaaggagcgcctacaagttgttaactactggaaaaa 960
DB 962 TTAATAATAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAA 1021
QY 961 tgtttgggcaaggattcagaattggcttctcaaaacagccactttggcagtgctact 1020
DB 1022 TGTTTGGGCAAGGATTCAGAAATTTGGCTCTCAAAACAGCCACTTTGGCAGTGCTACT 1081
QY 1021 tgggcagggttttaatgataccatgtaccatccgcctggaattggtgaatttgcagcc 1080
DB 1082 TGGCAGGTTTANTGATATCCATGTACCAATCCGCTGGAATGTGTGAATTTTGCTAGCC 1141
QY 1081 attgtctcatgaaccatctctgatttagcaaaagacgttaacagagtatcttaaaagtgggt 1140
DB 1142 ATTGTCTCATGAACCATCTCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTAGGT 1201
QY 1141 cacatgacctgaggagactattagacatgattgttattgtgcaatagttacagctgcta 1200

DB 1202 CACATGACCCGTGAGGAGCTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGCTA 1261
QY 1201 aaagagatatcttctggtcaatgatcactacttaatttctgagagagagacattag 1260
DB 1262 AAAAGGATATCTTCTGGTCAATGATCCTTACTTAAATTTTGTGAGAGAGAAACATTAG 1321
QY 1261 acaaacgatggagagtagtcgcaaaagaagccatgatgggacttgcaccaattataagaat 1320
DB 1322 ACAACGATGGAGAGTAGCGAAAGAGCCATGATGGGACTTGCCCAATTTATAAGAAAT 1381
QY 1321 atgctttacagtcagcagctggaaaaagatgctgcaaaacagatagcatgatacaaaaca 1380
DB 1382 ATGCTTTACAGTCAGCAGCTGGAAGAGATGCTCAAAACAGATAGCATGATCAAGACA 1441
QY 1381 aattgctacatatattatcaaaatagatttatgatcagctacttctgtgaacggatct 1440
DB 1442 AATTGCTACATATATATTATCAAAANTAGTATTGATGATCAGCTACTTGTGTAACGGATCT 1501
QY 1441 ttgctcaatacatggtgtccctccacaatttagaaactacagacacgagatgaatgcttatct 1500
DB 1502 TTGCTCAATACATGGTTCCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATT 1561
QY 1501 actgtatgccacacitggattttaaagtctgtaagcatitgaatgaaatgtgaaatgctc 1560
DB 1562 ACTGTATGCCACACCTGGATTAAATGCTGTGAAGCATTTGAATGAAATGTGAAATGTC 1621
QY 1561 aaaaactgctccgacacatcaagtaaaagatttgcctgacttgattaaagcaaccccaaacag 1620
DB 1622 AANAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGTATTAGCAACCCCAACAG 1681
QY 1621 atgccagtgccaaggccataatttcaaaagtgatggtgtattacaagaattttacctgctc 1680
DB 1682 ATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATC 1741
QY 1681 ctggttaaggtcaggatttcaatgaagaatttcacacaggtgttagaagaatgtagagaaaa 1740
DB 1742 CTGGTAAGGCTCAGGATTTTCATGAAGAAATTTACACAGGTGTTAGAAAGATGATGAGAAA 1801
QY 1741 taagaagcagtttagaagtagtcttctgtagtccacatgctcctcgaagcaggtgaaagtt 1800
DB 1802 TAGAANAAGCTTTAGAAAGTACTTTGTAGTCCACATGCTCTGCAAGCAGGCTGAAGGTT 1861
QY 1801 gtgtcgtgaaaataactaagaagttggggcaaccccaacacagcctacaaaatcccttctcgg 1860
DB 1862 GTGTGGCTGAAATAAATTAAGAAGTTGGGCAACCCCAACAGCCTACAAAATCCTTTCCTGG 1921
QY 1861 aagatgataagtttctctggagagatagcacctgtgcaatagatataccgaatctatca 1920
DB 1922 AAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGACATAGATACCGAATCTATCA 1981
QY 1921 gtgctcttataaacaagtgaacaaatcaatagatggaacagcagatgataagaatgaggt 1980
DB 1982 GTGCTCTTATTAAACAAGTGAACAATCAATAGATGAACAGCAGCATGATGAAGATGAGG 2041
QY 1981 gtgtccactgataagccatcagagcaggtcttgaactgcttaaggtactctctatcta 2040
DB 2042 GTGTTCCAACTGATCAAGCCATCAGAGCAGGCTTTGAACCTGCTTAAAGGTACTCTCATTTA 2101
QY 2041 cacatccctctcatttctctgctgaaacatttgaatcattactggtctgctgaaaa 2100
DB 2102 CACATCCCATCTCATTTTCATTTCTGCTGAACAATTTGAATCATTTACTTGGCTTGTCTGAAA 2161
QY 2101 tggatgatgaaaaagttagcagaagctgcactacaaattttcaaaaacacaggaagcaaaa 2160
DB 2162 TGGATCATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAA 2221
QY 2161 tgaagagagattttccacacatcagatcagccttgcctctgttttacaacacaaatcta 2220
DB 2222 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGGCTTCTGTTTACATCACAAATCTA 2281
QY 2221 aaaaagacccccctcaagccaaatagcatttctgataccatgcatgatttttcta 2280


```

exon      235..438
/number=3
exon      439..525
/number=4
exon      526..623
/number=5
exon      624..750
/number=6
exon      751..831
/number=7
exon      832..972
/number=8
exon      973..1088
/number=9
exon      1089..1183
/number=10
exon      1184..1329
/number=11
exon      1330..1481
/number=12
exon      1482..1595
/number=13
exon      1596..1677
/number=14
exon      1678..1726
/number=15
exon      1727..1866
/number=16
misc_feature 1866..2046
/number=17
misc_feature 1866..1941
/number=18
exon      1867..1982
/number=19
exon      1983..2088
/number=20
exon      2089..2249
/number=21
exon      2250..2373
/number=22
exon      2374..2532
/number=23
exon      2533..2601
/number=24
exon      2602..2738
/number=25
exon      2739..2862
/number=26
exon      2863..3067
/number=27
misc_feature 3068..3182
/number=28
exon      3183..3315
/number=29
misc_feature 3302..3572
/number=30
exon      3316..3435
/number=31
misc_feature 3349..3818
/number=32
exon      3819..4065
/number=33
exon      4066..4286
/number=34
misc_feature 4287..4419
/number=35
exon      4420..4674
/number=36
exon      4675..4921
/number=37
exon      4922..5168
/number=38
exon      5169..5415
/number=39
exon      5416..5662
/number=40
exon      5663..5909
/number=41
exon      5910..6156
/number=42
exon      6157..6403
/number=43
exon      6404..6650
/number=44
exon      6651..6897
/number=45
exon      6898..7144
/number=46
exon      7145..7391
/number=47
exon      7392..7638
/number=48
exon      7639..7885
/number=49
exon      7886..8132
/number=50
exon      8133..8379
/number=51
exon      8380..8626
/number=52
exon      8627..8873
/number=53
exon      8874..9120
/number=54
exon      9121..9367
/number=55
exon      9368..9614
/number=56
exon      9615..9861
/number=57
exon      9862..10108
/number=58
exon      10109..10355
/number=59
exon      10356..10602
/number=60
exon      10603..10849
/number=61
exon      10850..11096
/number=62
exon      11097..11343
/number=63
exon      11344..11590
/number=64
exon      11591..11837
/number=65
exon      11838..12084
/number=66
exon      12085..12331
/number=67
exon      12332..12578
/number=68
exon      12579..12825
/number=69
exon      12826..13072
/number=70
exon      13073..13319
/number=71
exon      13320..13566
/number=72
exon      13567..13813
/number=73
exon      13814..14060
/number=74
exon      14061..14307
/number=75
exon      14308..14554
/number=76
exon      14555..14801
/number=77
exon      14802..15048
/number=78
exon      15049..15295
/number=79
exon      15296..15542
/number=80
exon      15543..15789
/number=81
exon      15790..16036
/number=82
exon      16037..16283
/number=83
exon      16284..16530
/number=84
exon      16531..16777
/number=85
exon      16778..17024
/number=86
exon      17025..17271
/number=87
exon      17272..17518
/number=88
exon      17519..17765
/number=89
exon      17766..18012
/number=90
exon      18013..18259
/number=91
exon      18260..18506
/number=92
exon      18507..18753
/number=93
exon      18754..19000
/number=94
exon      19001..19247
/number=95
exon      19248..19494
/number=96
exon      19495..19741
/number=97
exon      19742..19988
/number=98
exon      19989..20235
/number=99
exon      20236..20482
/number=100
exon      20483..20729
/number=101
exon      20730..20976
/number=102
exon      20977..21223
/number=103
exon      21224..21470
/number=104
exon      21471..21717
/number=105
exon      21718..21964
/number=106
exon      21965..22211
/number=107
exon      22212..22458
/number=108
exon      22459..22705
/number=109
exon      22706..22952
/number=110
exon      22953..23199
/number=111
exon      23200..23446
/number=112
exon      23447..23693
/number=113
exon      23694..23940
/number=114
exon      23941..24187
/number=115
exon      24188..24434
/number=116
exon      24435..24681
/number=117
exon      24682..24928
/number=118
exon      24929..25175
/number=119
exon      25176..25422
/number=120
exon      25423..25669
/number=121
exon      25670..25916
/number=122
exon      25917..26163
/number=123
exon      26164..26410
/number=124
exon      26411..26657
/number=125
exon      26658..26904
/number=126
exon      26905..27151
/number=127
exon      27152..27398
/number=128
exon      27399..27645
/number=129
exon      27646..27892
/number=130
exon      27893..28139
/number=131
exon      28140..28386
/number=132
exon      28387..28633
/number=133
exon      28634..28880
/number=134
exon      28881..29127
/number=135
exon      29128..29374
/number=136
exon      29375..29621
/number=137
exon      29622..29868
/number=138
exon      29869..30115
/number=139
exon      30116..30362
/number=140
exon      30363..30609
/number=141
exon      30610..30856
/number=142
exon      30857..31103
/number=143
exon      31104..31350
/number=144
exon      31351..31597
/number=145
exon      31598..31844
/number=146
exon      31845..32090
/number=147
exon      32091..32337
/number=148
exon      32338..32584
/number=149
exon      32585..32831
/number=150
exon      32832..33078
/number=151
exon      33079..33325
/number=152
exon      33326..33572
/number=153
exon      33573..33819
/number=154
exon      33820..34066
/number=155
exon      34067..34313
/number=156
exon      34314..34560
/number=157
exon      34561..34807
/number=158
exon      34808..35054
/number=159
exon      35055..35301
/number=160
exon      35302..35548
/number=161
exon      35549..35795
/number=162
exon      35796..36042
/number=163
exon      36043..36289
/number=164
exon      36290..36536
/number=165
exon      36537..36783
/number=166
exon      36784..37030
/number=167
exon      37031..37277
/number=168
exon      37278..37524
/number=169
exon      37525..37771
/number=170
exon      37772..38018
/number=171
exon      38019..38265
/number=172
exon      38266..38512
/number=173
exon      38513..38759
/number=174
exon      38760..39006
/number=175
exon      39007..39253
/number=176
exon      39254..39500
/number=177
exon      39501..39747
/number=178
exon      39748..39994
/number=179
exon      40000..40246
/number=180
exon      40247..40493
/number=181
exon      40494..40740
/number=182
exon      40741..40987
/number=183
exon      40988..41234
/number=184
exon      41235..41481
/number=185
exon      41482..41728
/number=186
exon      41729..41975
/number=187
exon      41976..42222
/number=188
exon      42223..42469
/number=189
exon      42470..42716
/number=190
exon      42717..42963
/number=191
exon      42964..43210
/number=192
exon      43211..43457
/number=193
exon      43458..43704
/number=194
exon      43705..43951
/number=195
exon      43952..44198
/number=196
exon      44199..44445
/number=197
exon      44446..44692
/number=198
exon      44693..44939
/number=199
exon      44940..45186
/number=200
exon      45187..45433
/number=201
exon      45434..45680
/number=202
exon      45681..45927
/number=203
exon      45928..46174
/number=204
exon      46175..46421
/number=205
exon      46422..46668
/number=206
exon      46669..46915
/number=207
exon      46916..47162
/number=208
exon      47163..47409
/number=209
exon      47410..47656
/number=210
exon      47657..47903
/number=211
exon      47904..48150
/number=212
exon      48151..48397
/number=213
exon      48398..48644
/number=214
exon      48645..48891
/number=215
exon      48892..49138
/number=216
exon      49139..49385
/number=217
exon      49386..49632
/number=218
exon      49633..49879
/number=219
exon      49880..50126
/number=220
exon      50127..50373
/number=221
exon      50374..50620
/number=222
exon      50621..50867
/number=223
exon      50868..51114
/number=224
exon      51115..51361
/number=225
exon      51362..51608
/number=226
exon      51609..51855
/number=227
exon      51856..52102
/number=228
exon      52103..52349
/number=229
exon      52350..52596
/number=230
exon      52597..52843
/number=231
exon      52844..53090
/number=232
exon      53091..53337
/number=233
exon      53338..53584
/number=234
exon      53585..53831
/number=235
exon      53832..54078
/number=236
exon      54079..54325
/number=237
exon      54326..54572
/number=238
exon      54573..54819
/number=239
exon      54820..55066
/number=240
exon      55067..55313
/number=241
exon      55314..55560
/number=242
exon      55561..55807
/number=243
exon      55808..56054
/number=244
exon      56055..56301
/number=245
exon      56302..56548
/number=246
exon      56549..56795
/number=247
exon      56796..57042
/number=248
exon      57043..57289
/number=249
exon      57290..57536
/number=250
exon      57537..57783
/number=251
exon      57784..58030
/number=252
exon      58031..58277
/number=253
exon      58278..58524
/number=254
exon      58525..58771
/number=255
exon      58772..59018
/number=256
exon      59019..59265
/number=257
exon      59266..59512
/number=258
exon      59513..59759
/number=259
exon      59760..60006
/number=260
exon      60007..60253
/number=261
exon      60254..60500
/number=262
exon      60501..60747
/number=263
exon      60748..60994
/number=264
exon      60995..61241
/number=265
exon      61242..61488
/number=266
exon      61489..61735
/number=267
exon      61736..61982
/number=268
exon      61983..62229
/number=269
exon      62230..62476
/number=270
exon      62477..62723
/number=271
exon      62724..62970
/number=272
exon      62971..63217
/number=273
exon      63218..63464
/number=274
exon      63465..63710
/number=275
exon      63711..63957
/number=276
exon      63958..64204
/number=277
exon      64205..64451
/number=278
exon      64452..64698
/number=279
exon      64699..64945
/number=280
exon      64946..65192
/number=281
exon      65193..65439
/number=282
exon      65440..65686
/number=283
exon      65687..65933
/number=284
exon      65934..66179
/number=285
exon      66180..66426
/number=286
exon      66427..66673
/number=287
exon      66674..66920
/number=288
exon      66921..67167
/number=289
exon      67168..67414
/number=290
exon      67415..67661
/number=291
exon      67662..67908
/number=292
exon      67909..68155
/number=293
exon      68156..68402
/number=294
exon      68403..68649
/number=295
exon      68650..68896
/number=296
exon      68897..69143
/number=297
exon      69144..69390
/number=298
exon      69391..69637
/number=299
exon      69638..69884
/number=300
exon      69885..70131
/number=301
exon      70132..70378
/number=302
exon      70379..70625
/number=303
exon      70626..70872
/number=304
exon      70873..71119
/number=305
exon      71120..71366
/number=306
exon      71367..71613
/number=307
exon      71614..71860
/number=308
exon      71861..72107
/number=309
exon      72108..72354
/number=310
exon      72355..72601
/number=311
exon      72602..72848
/number=312
exon      72849..73095
/number=313
exon      73096..73342
/number=314
exon      73343..73589
/number=315
exon      73590..73836
/number=316
exon      73837..74083
/number=317
exon      74084..74330
/number=318
exon      74331..74577
/number=319
exon      74578..74824
/number=320
exon      74825..75071
/number=321
exon      75072..75318
/number=322
exon      75319..75565
/number=323
exon      75566..75812
/number=324
exon      75813..76059
/number=325
exon      76060..76306
/number=326
exon      76307..76553
/number=327
exon      76554..76800
/number=328
exon      76801..77047
/number=329
exon      77048..77294
/number=330
exon      77295..77541
/number=331
exon      77542..77788
/number=332
exon      77789..78035
/number=333
exon      78036..78282
/number=334
exon      78283..78529
/number=335
exon      78530..78776
/number=336
exon      78777..79023
/number=337
exon      79024..79270
/number=338
exon      79271..79517
/number=339
exon      79518..79764
/number=340
exon      79765..80011
/number=341
exon      80012..80258
/number=342
exon      80259..80505
/number=343
exon      80506..80752
/number=344
exon      80753..81000
/number=345
exon      81001..81246
/number=346
exon      81247..81493
/number=347
exon      81494..81740
/number=348
exon      81741..81987
/number=349
exon      81988..82234
/number=350
exon      82235..82481
/number=351
exon      82482..82728
/number=352
exon      82729..82975
/number=353
exon      82976..83222
/number=354
exon      83223..83469
/number=355
exon      83470..83716
/number=356
exon      83717..83963
/number=357
exon      83964..84210
/number=358
exon      84211..84457
/number=359
exon      84458..84704
/number=360
exon      84705..84951
/number=361
exon      84952..85198
/number=362
exon      85199..85445
/number=363
exon      85446..85692
/number=364
exon      85693..85939
/number=365
exon      85940..86186
/number=366
exon      86187..86433
/number=367
exon      86434..86680
/number=368
exon      86681..86927
/number=369
exon      86928..87174
/number=370
exon      87175..87421
/number=371
exon      87422..87668
/number=372
exon      87669..87915
/number=373
exon      87916..88162
/number=374
exon      88163..88409
/number=375
exon      88410..88656
/number=376
exon      88657..88903
/number=377
exon      88904..89150
/number=378
exon      89151..89397
/number=379
exon      89398..89644
/number=380
exon      89645..89891
/number=381
exon      89892..90138
/number=382
exon      90139..90385
/number=383
exon      90386..90632
/number=384
exon      90633..90879
/number=385
exon      90880..91126
/number=386
exon      91127..91373
/number=387
exon      91374..91620
/number=388
exon      91621..91867
/number=389
exon      91868..92114
/number=390
exon      92115..92361
/number=391
exon      92362..92608
/number=392
exon      92609..92855
/number=393
exon      92856..93102
/number=394
exon      93103..93349
/number=395
exon      93350..93596
/number=396
exon      93597..93843
/number=397
exon      93844..94090
/number=398
exon      94091..94337
/number=399
exon      94338..94584
/number=400
exon      94585..94831
/number=401
exon      94832..95078
/number=402
exon      95079..95325
/number=403
exon      95326..95572
/number=404
exon      95573..95819
/number=405
exon      95820..96066
/number=406
exon      96067..96313
/number=407
exon      96314..96560
/number=408
exon      96561..96807
/number=409
exon      96808..97054
/number=410
exon      97055..97301
/number=411
exon      97302..97548
/number=412
exon      97549..97795
/number=413
exon      97796..98042
/number=414
exon      98043..98288
/number=415
exon      98289..98535
/number=416
exon      98536..98782
/number=417
exon      98783..99029
/number=418
exon      99030..99276
/number=419
exon      99277..99523
/number=420
exon      99524..99770
/number=421
exon      99771..100017
/number=422
exon      100018..100264
/number=423
exon      100265..100511
/number=424
exon      100512..100758
/number=425
exon      100759..101005
/number=426
exon      101006..101252
/number=427
exon      101253..101499
/number=428
exon      101500..101746
/number=429
exon      101747..101993
/number=430
exon      101994..102240
/number=431
exon      102241..102487
/number=432
exon      102488..102734
/number=433
exon      102735..102981
/number=434
exon      102982..103228
/number=435
exon      103229..103475
/number=436
exon      103476..103722
/number=437
exon      103723..103969
/number=438
exon      103970..104216
/number=439
exon      104217..104463
/number=440
exon      104464..104710
/number=441
exon      104711..104957
/number=442
exon      104958..105204
/number=443
exon      105205..105451
/number=444
exon      105452..105698
/number=445
exon      105699..105945
/number=446
exon      105946..106192
/number=447
exon      106193..106439
/number=448
exon      106440..106686
/number=449
exon      106687..106933
/number=450
exon      106934..107179
/number=451
exon      107180..107426
/number=452
exon      107427..107673
/number=453
exon      107674..107920
/number=454
exon      107921..108167
/number=455
exon      108168..108414
/number=456
exon      108415..108661
/number=457
exon      108662..108908
/number=458
exon      108909..109155
/number=459
exon      109156..109402
/number=460
exon      109403..109649
/number=461
exon      109650..109896
/number=462
exon      109897..110143
/number=463
exon      110144..110389
/number=464
exon      110390..110636
/number=465
exon      110637..110883
/number=466
exon      110884..111130
/number=467
exon      111131..111377
/number=468
exon      111378..111624
/number=469
exon      111625..111871
/number=470
exon      111872..112118
/number=471
exon      112119..112365
/number=472
exon      112366..112612
/number=473
exon      112613..112859
/number=474
exon      112860..113106
/number=475
exon      113107..113353
/number=476
exon      113354..113599
/number=477
exon      113600..113846
/number=478
exon      113847..114093
/number=479
exon      114094..114339
/number=480
exon      114340..114586
/number=481
exon      114587..114833
/number=482
exon      114834..115079
/number=483
exon      115080..115326
/number=484
exon      115327..115572
/number=485
exon      115573..115819
/number=486
exon      115820..116066
/number=487
exon      116067..116312
/number=488
exon      116313..116558
/number=489
exon      116559..116805
/number=490
exon      116806..117051
/number=491
exon      117052..117297
/number=492
exon      117298..117544
/number=493
exon      117545..117790
/number=494
exon      117791..118036
/number=495
exon      118037..118283
/number=496
exon      118284..118529
/number=497
exon      118530..118775
/number=498
exon      118776..119021
/number=499
exon      119022..119268
/number=500
exon      119269..119515
/number=501
exon      119516..119762
/number=502
exon      119763..120009
/number=503
exon      120010..120256
/number=504
exon      120257..120503
/number=505
exon      120504..120749
/number=506
exon      120750..121000
/number=507
exon      121001..121246
/number=508
exon      121247..121492
/number=509
exon      121493..121739
/number=510
exon      121740..121986
/number=511
exon      121987..122232
/number=512
exon      122233..122478
/number=513
exon      122479..122725
/number=514
exon      122726..122971
/number=515
exon      122972..123217
/number=516
exon      123218..123463
/number=517
exon      123464..123710
/number=518
exon      123711..123956
/number=519
exon      123957..124203
/number=520
exon      124204..124449
/number=521
exon      124450..124696
/number=522
exon      124697..124942
/number=523
exon      124943..125188
/number=524
exon      125189..125435
/number=525
exon      125436..125681
/number=526
exon      125682..125927
/number=527
exon      125928..126173
/number=528
exon      126174..126419
/number=529
exon      126420..126666
/number=530
exon      126667..126912
/number=531
exon      126913..127158
/number=532
exon      127159..127405
/number=533
exon      127406..127651
/number=534
```


1861 AAATGATCAAGTTTCTTCTTGAGAGGATAGCACCTGTCACATAGATACCGAATCTATCA 1920
1921 gtgctcttattaaacagtgtaacaaatacaatagatgaacagagatgaagataggg 1980
1921 GTGCTCTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGG 1980
1981 gtgttccaactgatacgaagccatcagacagagctcttgtaactgcttaaggtactctcattta 2040
1981 GTGTTCCAATGATCAAGCCATCAGACAGAGCTCTTGAACCTTGAAGGTACTCTCATTTA 2040
2041 caactcccatctcatttcattctgctgtaaacatttggaatcattactgctgctgtaaaa 2100
2041 CACATCCCATCTCATTTCAATCTGCTGAAACATTTGAATCATTTACTGCGCTGTCTGAAA 2100
2101 tgaatgataaaagtagcagaagctcactacaataatttcaaaaacacaggaagcaaaa 2160
2101 TGGATGATGAAAAGTAGCAGAAGCTGCACCTACAAAATTTTCAAAAACACAGGAAGCAAAA 2160
2161 ttgaagaggattttccacacatcagatcagccttgcttctcctggttttacatcacaaatcta 2220
2161 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGGCTTCCGTGTTTACATCACAAATCTA 2220
2221 aaaaagaccccccgctgaagccaaaatgcccatttcattgtatccatgcgatatatttcta 2280
2221 AAAAAGGACCCCCCGTCAAGCCAAATATGCCATTCATTTGTATCCATGCGATATTTTCTA 2280
2281 gtaagagacccatttgacagatatttgagcctctgcataaagagcctagatccaaagca 2340
2281 GTAAAGAGGCCGATTTGACAGATATTTGAGCCTCTGCAATAAGAGGCTAGATCAAGCA 2340
2341 acctggaacatctcataacaccattggttactattggtcatattgctctcctctgcacctg 2400
2341 ACCTTGGAACTCTCATTAACACCATTTGTTACTATTGGTCATATTGCTCTCTTGCACCTG 2400
2401 atcaatttgctgctccttggaatctctgggtagtaatttcatttgtaagagattcttca 2460
2401 ATCAATTTGCTGCTTCTTGAATCTTTGTTAGCTTCTTCAATTTGTTGAAGATCTTCTCA 2460
2461 tgaatgacgcttccagggaagaagcaactaaacttgggttccagatgaagaagtat 2520
2461 TGAATGATCGGCTTCCAGGGAAGAAGCACTAAACTTTTGGGTTTCCAGATGAAGAAGTAT 2520
2521 ctcttgagacaattggtcaaaattcaggctattaaaaatgagttcgtggtactttgaa 2580
2521 CTCCTGAGACAAATGGTCAAAATTCAGGCTATTAAATGATGGTTTCGATGGCTACTTGAA 2580
2581 tgaataataatcagtaaatcaggaactctacctaagaattgctaaagattgctaaacaatatgc 2640
2581 TGAAAAATAATCAGATTAATTCAGGAATCTTACCTTTAAGATTGCTTAACAAATATTGC 2640
2641 atagtatgagacttgacagaacaggggaaaattagtaaacagatatgtcacgtctga 2700
2641 ATAGTATGGAGACTTGACAGAACAGGGGAAAATTAGTAACACAGATGTGTCAGCTCTGA 2700
2701 gacttgcctggtgagtgctatttgaaagctggcacagaacacctgttaccatgaaatca 2760
2701 GACTTGTCTGGAGTGTCTATTGTGAAGCTGGCACAGAACCCTGTTTACCATGAATCA 2760
2761 tcacattagaacaataatcagctatgtcattagctaatacagatgaatgctatcaagtaa 2820
2761 TCACATTAGAACAATATCAGCTATGTGCATTTAGCTATCAACAGATGAATGCTTATCAAGTAA 2820
2821 gacaagtgttggccagaacattcacaaagcctttcccgtttacgcttccacttaagt 2880
2821 GACAAGTGTTTGGCCAGAAACTTCACAAAGGCCCTTTCCCGTTTACGCTTCCACTTGTAGT 2880
2881 atatggcaattctgcccctttgtgcaaaagatcctgttaaaagagagaagagctcatgcta 2940
2881 ATATGGCAATCTGTGCCCTTTGTCAAAAGATCTGTAAAGSAGAGAAGACCTCATGCTA 2940
2941 ggcgaattgttgtaaaaaataataatgtaagcg9gg9agtaactgaagcagcagatgcagctg 3000
2941 GGCAATGTTTGGTGAATAATAATGTAAGCGGGAGTATCTGAAGCAGCATGCAGCTG 3000

QY 3001 ttagtgaataattatgtctcttaccagagtattgtgttctccatatacaattccacctt 3060
Db 3001 TTAGTGAAAAATYATTGTCTCTTCTACAGAGTATGTTGTTCCATATACAATTCACCTTT 3060
QY 3061 tggcacatgacccagattatgtcaaaagtacaggatatgaacaacttaaaagtattaaag 3120
Db 3061 TGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACATTAAAGATGTTAAAG 3120
QY 3121 aatgtctttgttctctggaaatattaatgtgctaaaaatgaaaaatacaagtcacgctt 3180
Db 3121 AATGCTTTGTTGTTCTTGAAATATTAAATGGCTAAAAATGAAATTAACAGTACCGCTT 3180
QY 3181 ttatcagaagaatgttagaaaaatttaaaacaaacaaagatgcccaagacacagatgatg 3240
Db 3181 TTATCAGAAAGATGGTAGAAAATATTAAAGCAAAAGATGCCCAAGGACCAAGATGATG 3240
QY 3241 caaaaatgaatgaaaaactgtacactgtgtgtgattgtgcatgaataatcatcatgctcaa 3300
Db 3241 CAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAA 3300
QY 3301 agagtactaatacagatttggaatctctctaaagaccgcggtactaccagctcgtttcttca 3360
Db 3301 AGAGTACTACATACAGTTTGGAAATCTCTTAAAGACCCCGGTACTACCAGCTCGTTCTTCA 3360
QY 3361 ctcaacctgacaagaatttcagtaaccacaaaattatctctcctcctgaaatgaatcat 3420
Db 3361 CTCAACTGTACAGAAATTTTCAAGTAAACCAAAATATTCTGCCCTCTGGAATGAATCAT 3420
QY 3421 ttttctcctctgaaaaaccttaaaacaaacaaatgttctaggagctgttaacaagccacttt 3480
Db 3421 TTTTCACTCTCTGAAAAACCTTAAACAAACCAATGTTCTAGGAGCTGTTAAACAAGCCACTTT 3480
QY 3481 catcagcaggcaagcaatctcagaccaaaatcatcacgaatggaactgtaagcaatgcaa 3540
Db 3481 CATCAGCAGGCAAGCAATCTCAGACCAAAATCATCACGAATGGAACACTGTAGCAATGCAA 3540
QY 3541 qcacagctcaaaatcccaagctctcctgagaataaaagggagcttgatagttcttgaaa 3600
Db 3541 GCACAGCTCAAAATCCAAAGCTCTCTGGAAGATAAAGGGGAGGCTTGATAGTTCTGAAA 3600
QY 3601 tggatcagtgaaaaatgaagattacacaaatgtcttccacctttccgggggaaaaaagt 3660
Db 3601 TGGATCAGTGAAAAATGAAGATTACAAATGTCTTCACTTTTCCGGGGAAAAAGAGATG 3660
QY 3661 acaagagagacgactctgacttcttaaggtctgaattggaagaaaccttagagggcagaaaa 3720
Db 3661 ACAAGAGAGAGAGCTCTGATCTTGTAAAGTCTGAATTTGGAGAAGCCTTAGAGGCAGAAAA 3720
QY 3721 aaacgcccgtcacagaacagagagaaattagtatggtatgacttgactaagttggttac 3780
Db 3721 AAACCCCGTCCACAGAACGGGAGGAGAAATTTAGGTATGGATGACTTGACTAAGTTGGTAC 3780
QY 3781 aggaacgaacacctaaaggcagtcagcgaagtgcgaaaaagggccacacacccctctctgagt 3840
Db 3781 AGGAACAGAGACCTTAAAGGCAGTCAGCGAAGTCCGAAAAAGAGCCATACGCTTTCAGAA 3840
QY 3841 ctgattgaacagcagtgccctgaggaagaagaggtctcaaaagaagatatattagaagaatgaag 3900
Db 3841 CTGATGAACACAGCTGCGCTCAGGAAAAAGAGGCTCAAAAGAAGATATATTAGAANAATGAAG 3900
QY 3901 atgaacagaaatagtcgccccaaaaagggtaaaagagggcgcgaccaccaaaacctctctgtg 3960
Db 3901 ATGAACAGATAGTCCCGCCAAAAAAGGTTAAAGAGCCGACCAACCAAACTCTTGTGTG 3960
QY 3961 gaggtacacaaaaagaagagccaaatgaaaacttctaaaaaaggaagcaaaaaaaat 4020
Db 3961 GAGGTACACCAAAAAGAGAGGCCAAATGAAAACCTTCTTAAAAAAGGAAGCAAAAAAAT 4020
QY 4021 ctggaacctccagaccagagagagaggaagaagaagaagcaaaagtggaaataacggaac 4080
Db 4021 CTGGACCTCCAGCACCCAGAGGAGGAGGAAGAAGAAAGAAAGCAAAAGTGAATACGGAAC 4080

TITLE Direct Submission
JOURNAL Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REFERENCE 4 (bases 1 to 5253)
AUTHORS Geck, P., Szelele, J., Jimenez, J., Sonnenschein, C. and Soto, A. M.
JOURNAL Direct Submission
Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REMARK Sequence update by submitter
COMMENT On Apr 5, 1999 this sequence version replaced gi:4539617.
FEATURES
 Location/Qualifiers
 1..5253
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="13q12-q13"
 /cell_line="LNCaP"
 /tissue_type="cancer cells"
 /notes="exon 1 maps to p1 Artificial Chromosome PAC26H23;
 exons 2-24 map to cosmid 267p19; exons 25-34 map to p1
 Artificial Chromosome PAC 49010"
 1..5253
 /gene="AS3"
 1..46
 /gene="AS3"
 /number=1
 47..173
 /gene="AS3"
 /number=2
 66..4241
 /gene="AS3"
 /codon_start=1
 /product="androgen-induced prostate proliferative shutoff
 associated protein"
 /protein_id="AAD22134.2"
 /db_xref="GI:4559410"
 /translation="WAHSKTRTNDGKTYPPGVKTEISDKSEEMVRLKMYVKTQFMD
 MQDSEERKELYNLALHLSADFFLKHDPKDVRLVACCLADIERLYAPEAPYTPDK
 LADIPNFTIRQLKGLDTSKSPQFNRYFYLENNIWNKSYNICFELEDSNEITQLYRT
 LFSVINNGHQVHHMVDLMSIIICEGDTVSQELLDVTLVNLVPAHKNLKOAYDLA
 KALLKRTAQAIIEPYITNFFNQVLMGKTSISLSEHVSLLILELYNIDSHLLSLVLPQ
 LEFLKSDNEERLQVLLKAKFGAKDSELSQNKPLWCQYLGKRFNDIHTVIRLECV
 KFAHCLMNHDPDLAKTEYLKVRSHDPEAIRHDVIVSIVTAARKDILLVNDHLNF
 VGERITDKWRVRKEMMGIAOIYKKYALQSAAGDAAKOIAWIKDKLHLYQNSID
 DRLLEIRIQAQWVPHNLETTERMKCLYLYATLDLNAVKNALNEMKCONLRRHQVKD
 LDDLINQDASVRAIFSKVMVITKSLPDPKQDMKKFTQVLEDDKIRKQLEVL
 VSPKSCQAEGCVREITKKNPKOPTNPFLEIMFKFLIERIAPVHIDTESISALIKQ
 VNSIDGTADDEGVPTDQAIKAGLELLKLSFTHPISFSAETPFESLACIKMDEE
 KVAEALQIFKNTGSKIIEEDFPHISALLPVLLHKSCKGPPQAKYAICHIAIFSSK
 EAQAFIIFPLHKSLLDPSNLEHILPTLVIGHALLAPQFAAPLKSIVATFIVKDLL
 MDRLPGKTKTLKLVDPDEVSPETWVKIQIKMMVRLWLLGMNNHNSKSTSTLRLLT
 ILHSDGLTFQSKISKSPDMSRLRLAAGSAIVKADQPCYHEITLEQYQCALAINDE
 CYQVRQVFAQLKHLKSLRLPLEYMAICALKADPKERRAHARQCLVKNTNVRREY
 LKQHNASHSEKLSLLPEYVVPYVITHLHADPDYVKVDIEQLKQVKECLWFVLEILMA
 KNENSHAFIRKVENIKOTKDAQGDDAKMEKLYTCDVAMNIIIMSKSTYISLESP
 KDPVLPAFFTOPDKNFNTKNYLPFGKMSFPTPGPKPTNVLGAVNKPFLSAGQSQ
 TKSSRMETVGNASSSNPSPGRIKGRDSSMDHSENEEDYTMSPLPKGRDKRDDS
 DIVRSELEKPRKPTVTEREELKGMDDTLKVDQRPKQSRKRGHTASDEQ
 QWPEERKLEDILENEDEQSPKKGKRPKPLGGGTPEEPTMTKTSKGSKKSG
 PPAPEEEEEERQSGNTEQSKSKQHRVSRRAQRAESSESAIESTQSTPQKGRGP
 SKTPSPQPKKNV"
 174..377
 /gene="AS3"
 378..464
 /number=3
 /gene="AS3"
 465..562
 /number=4
 /gene="AS3"
 563..689
 /number=5
 /gene="AS3"
 690..770
 /number=6
 /gene="AS3"
 771..911
 /number=7
 /gene="AS3"
 912..1027
 /number=8
 /gene="AS3"
 1028..1122
 /number=9
 /gene="AS3"
 1123..1268
 /number=10
 /gene="AS3"
 1269..1420
 /number=11
 /gene="AS3"
 1421..1534
 /number=12
 /gene="AS3"
 1535..1616
 /number=13
 /gene="AS3"
 1617..1665
 /number=14
 /gene="AS3"
 1666..1805
 /number=15
 /gene="AS3"
 1806..1921
 /number=16
 /gene="AS3"
 1922..2027
 /number=17
 /gene="AS3"
 2028..2188
 /number=18
 /gene="AS3"
 2189..2312
 /number=19
 /gene="AS3"
 2313..2471
 /number=20
 /gene="AS3"
 2472..2540
 /number=21
 /gene="AS3"
 2541..2677
 /number=22
 /gene="AS3"
 2678..2801
 /number=23
 /gene="AS3"
 2802..3006
 /number=24
 /gene="AS3"
 3007..3121
 /number=25
 /gene="AS3"
 3122..3254
 /number=26
 /gene="AS3"
 3255..3374
 /number=27
 /gene="AS3"
 3375..3437
 /number=28
 /gene="AS3"
 3438..3583
 /number=29
 /gene="AS3"
 3584..3770
 /number=30

885	27	0.5	69282	71	AC040912	Homo sapi	AC040912	Homo sapi	83	AP002079	Homo sapi
886	27	0.5	69290	76	AC084206	Homo sapi	AC084206	Homo sapi	68	AC024051	Homo sapi
887	27	0.5	69579	72	AC055831	Homo sapi	AC055831	Homo sapi	85	AC002558	Homo sapi
888	27	0.5	70105	67	AC007732	Homo sapi	AC007732	Homo sapi	78	AL138935	Homo sapi
889	27	0.5	70944	60	AC022723	Homo sapi	AC022723	Homo sapi	91	AP000692	Homo sapi
890	27	0.5	71118	82	AP000586	Homo sapi	AP000586	Homo sapi	90	AL359813	Human DNA
891	27	0.5	71983	80	AL445429	Human DNA	AL445429	Human DNA	92	HS187N21	Human DNA
892	27	0.5	73029	78	AC090560	Homo sapi	AC090560	Homo sapi	85	AC003087	Human DNA
893	27	0.5	73037	77	AC090247	Homo sapi	AC090247	Homo sapi	85	AC004836	Homo sapi
894	27	0.5	73087	92	HS147M19	Homo sapi	HS147M19	Homo sapi	84	HS147M19	Homo sapi
895	27	0.5	73607	90	AL162740	Human DNA	AL162740	Human DNA	82	AL590433	Mus muscu
896	27	0.5	73701	64	AC016402	Homo sapi	AC016402	Homo sapi	77	AC087771	Medicago
897	27	0.5	73754	82	AP000592	Homo sapi	AP000592	Homo sapi	92	HS1185H19	Human DNA
898	27	0.5	74989	67	AC023333	Homo sapi	AC023333	Homo sapi	93	AL327773	Human DNA
899	27	0.5	75582	89	AL138729	Human DNA	AL138729	Human DNA	90	AL355272	Human DNA
900	27	0.5	75863	64	AC016059	Homo sapi	AC016059	Homo sapi	92	HS106120B	Human DNA
901	27	0.5	76086	77	AC084246	Homo sapi	AC084246	Homo sapi	86	AC005629	Homo sapi
902	27	0.5	77498	90	AL162505	Human DNA	AL162505	Human DNA	87	AC011450	Homo sapi
903	27	0.5	78404	90	AL445198	Human DNA	AL445198	Human DNA	85	AC005102	Homo sapi
904	27	0.5	78405	73	AC068094	Homo sapi	AC068094	Homo sapi	94	AF130357	Mus muscu
905	27	0.5	78692	65	AC018370	Homo sapi	AC018370	Homo sapi	85	AC004929	Homo sapi
906	27	0.5	80920	85	AC002036	Homo sapi	AC002036	Homo sapi	79	AL157767	Homo sapi
907	27	0.5	81524	93	HS147M19	Homo sapi	HS147M19	Homo sapi	85	AC004263	Homo sapi
908	27	0.5	81524	93	HS147M19	Homo sapi	HS147M19	Homo sapi	85	AC001226	Genomic S
909	27	0.5	82359	85	AC004922	Homo sapi	AC004922	Homo sapi	91	AP002078	Homo sapi
910	27	0.5	83059	91	AP001435	Homo sapi	AP001435	Homo sapi	87	AC011450	Homo sapi
911	27	0.5	84300	93	HS104A17	Human DNA	HS104A17	Human DNA	77	AC084697	Homo sapi
912	27	0.5	84371	83	AP001963	Homo sapi	AP001963	Homo sapi	90	AL139042	Human DNA
913	27	0.5	84478	89	AL133462	Human DNA	AL133462	Human DNA	67	AC022470	Homo sapi
914	27	0.5	85565	89	AL136133	Human DNA	AL136133	Human DNA	90	AL353596	Human DNA
915	27	0.5	85624	90	AL133933	Human DNA	AL133933	Human DNA	85	AC000403	Genomic S
916	27	0.5	85952	93	HS147M19	Human DNA	HS147M19	Human DNA	60	AC008723	Homo sapi
917	27	0.5	86113	89	AF285442	Homo sapi	AF285442	Homo sapi	14	F23A5	
918	27	0.5	86437	77	AC090276	Homo sapi	AC090276	Homo sapi	90	AL390208	Human DNA
919	27	0.5	86437	77	AC090276	Homo sapi	AC090276	Homo sapi	60	AC036565	Continuation (6 of
920	27	0.5	86766	83	AP001872	Homo sapi	AP001872	Homo sapi	70	AC085762	Continuation (3 of
921	27	0.5	87022	82	AP001089	Homo sapi	AP001089	Homo sapi	70	AC026388	Continuation (2 of
922	27	0.5	87789	89	AL133270	Human DNA	AL133270	Human DNA	76	AC079517	Continuation (4 of
923	27	0.5	87844	14	BN245480	Brassica	BN245480	Brassica	92	HS281H8	Human DNA
924	27	0.5	88036	97	AC021487	Human Chr	AC021487	Human Chr	92	HS191E19	Human DNA
925	27	0.5	88309	66	AC008458	Homo sapi	AC008458	Homo sapi	82	AP000407	Homo sapi
926	27	0.5	89015	60	AC008458	Homo sapi	AC008458	Homo sapi	86	AC006375	Homo sapi
927	27	0.5	89251	93	HS147M19	Human DNA	HS147M19	Human DNA	86	AC006486	Homo sapi
928	27	0.5	89292	91	AP001252	Homo sapi	AP001252	Homo sapi	86	AC006486	Homo sapi
929	27	0.5	89743	94	AC003062	Mouse Chr	AC003062	Mouse Chr	92	HS15619	Homo sapi
930	27	0.5	90605	70	AC026450	Homo sapi	AC026450	Homo sapi			
931	27	0.5	90650	94	AF242432	Mus muscu	AF242432	Mus muscu			
932	27	0.5	91202	60	AC006186	Homo sapi	AC006186	Homo sapi			
933	27	0.5	92569	90	AL353688	Human DNA	AL353688	Human DNA			
934	27	0.5	93151	77	AC090117	Homo sapi	AC090117	Homo sapi			
935	27	0.5	93223	66	AC021825	Homo sapi	AC021825	Homo sapi			
936	27	0.5	94212	91	AP000246	Homo sapi	AP000246	Homo sapi			
937	27	0.5	95360	67	AC022465	Homo sapi	AC022465	Homo sapi			
938	27	0.5	95369	90	AL359694	Human DNA	AL359694	Human DNA			
939	27	0.5	96489	12	AC009894	Arabidops	AC009894	Arabidops			
940	27	0.5	96500	91	AP001243	Homo sapi	AP001243	Homo sapi			
941	27	0.5	97799	79	AL354890	Human DNA	AL354890	Human DNA			
942	27	0.5	97835	92	HS292E10	Human DNA	HS292E10	Human DNA			
943	27	0.5	98543	85	AC005352	Homo sapi	AC005352	Homo sapi			
944	27	0.5	98562	87	AC008969	Homo sapi	AC008969	Homo sapi			
945	27	0.5	99332	93	HS1298J18	Human DNA	HS1298J18	Human DNA			
946	27	0.5	99342	92	HS201D17	Human DNA	HS201D17	Human DNA			
947	27	0.5	99344	86	AC006271	Homo sapi	AC006271	Homo sapi			
948	27	0.5	99370	85	AC005057	Homo sapi	AC005057	Homo sapi			
949	27	0.5	99716	91	AL512454	Human DNA	AL512454	Human DNA			
950	27	0.5	99908	86	AC006464	Homo sapi	AC006464	Homo sapi			
951	27	0.5	100000	91	AP000130	Homo sapi	AP000130	Homo sapi			
952	27	0.5	100000	91	AP000208	Homo sapi	AP000208	Homo sapi			
953	27	0.5	100139	88	AC021752	Human DNA	AC021752	Human DNA			
954	27	0.5	100272	93	HS147M19	Human DNA	HS147M19	Human DNA			
955	27	0.5	100627	86	AC007359	Homo sapi	AC007359	Homo sapi			
956	27	0.5	100953	87	AC009489	Homo sapi	AC009489	Homo sapi			
957	27	0.5	101006	76	AC079392	Homo sapi	AC079392	Homo sapi			

ALIGNMENTS

RESULT 1

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

HS095825

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

c 739	28	0.5	223398	78	AL137161	AL137161 Homo sapi	812	27	0.5	36767	85	AC004646	AC004646 Homo sapi
c 740	28	0.5	225300	77	AC084798	AC084798 Mus muscu	813	27	0.5	37170	93	HSE129H9	268224 Human DNA s
c 741	28	0.5	225371	81	AL445222	AL445222 Homo sapi	814	27	0.5	37655	86	AC007792	AC007792 Homo sapi
c 742	28	0.5	226841	84	HSAC002043	AC002043 Homo sapi	815	27	0.5	37930	86	AC005564	AC005564 Homo sapi
c 743	28	0.5	227008	82	AF001771	AP001771 Homo sapi	c 816	27	0.5	38258	85	AC000402	AC000402 Genomic s
c 744	28	0.5	229962	76	AC0079520	AP0079520 Mus muscu	c 817	27	0.5	38682	93	HSAC002077	AC002077 Human cos
c 745	28	0.5	237678	94	AC055818	AC055818 Mus muscu	c 818	27	0.5	39595	85	AC004509	AC004509 Homo sapi
c 746	28	0.5	249687	63	AC015693	AC015693 Homo sapi	c 819	27	0.5	39854	88	AC020954	AC020954 Homo sapi
c 747	28	0.5	252370	4	AE003529	AE003529 Drosophil	c 820	27	0.5	40582	91	AP000691	AP000691 Homo sapi
c 748	28	0.5	267685	68	AC024901	AC024901 Homo sapi	821	27	0.5	40839	87	AC026805	AC026805 Homo sapi
c 749	28	0.5	283512	80	AL365202	AL365202 Homo sapi	822	27	0.5	40709	97	L48038	AC026805 Homo sapi
c 750	28	0.5	300000	91	AF002530	AP002530 Homo sapi	c 823	27	0.5	41015	92	HS590K14	L48038 Structure o
c 751	28	0.5	300691	60	AC005842	AC005842 Homo sapi	c 824	27	0.5	41052	85	AC002499	AL049198 Human DNA
c 752	28	0.5	321193	70	AC026340	AC026340 Homo sapi	c 825	27	0.5	41100	91	AP000540	AC002499 Human Cos
c 753	28	0.5	336880	66	AC021052	AC021052 Homo sapi	c 826	27	0.5	41275	85	AC004182	AP000540 Homo sapi
c 754	28	0.5	340000	91	AP001724	AP001724 Homo sapi	c 827	27	0.5	41277	85	AC004180	AC004182 Homo sapi
c 755	27	0.5	213	14	SCB307346	AP001724 Homo sapi	c 828	27	0.5	41308	85	AC005203	AC004180 Homo sapi
c 756	27	0.5	213	15	TAE307350	AJ307350 Arabidops	c 829	27	0.5	41613	85	AC004790	AC005203 Homo sapi
c 757	27	0.5	306	88	AF103994	AJ307350 Arabidops	c 830	27	0.5	41696	86	AC005932	AC004790 Homo sapi
c 758	27	0.5	353	97	HUMCOL1A01	AF103994 Cebus ape	c 831	27	0.5	41766	90	AL359555	AC005932 Homo sapi
c 759	27	0.5	359	97	HUMCOL1A01	M32790 Human alpha	c 832	27	0.5	42307	87	AC010510	AL359555 Human DNA
c 760	27	0.5	390	13	ATH307357	M32798 Human alpha	c 833	27	0.5	42569	89	AL138915	AC010510 Homo sapi
c 761	27	0.5	390	13	ATH307358	AJ307357 Arabidops	c 834	27	0.5	42854	91	AP001242	AL138915 Human DNA
c 762	27	0.5	392	13	ATH307359	AJ307358 Arabidops	c 835	27	0.5	44078	85	AC002102	AP001242 Homo sapi
c 763	27	0.5	428	96	PWNSLRNB	AJ307359 Arabidops	c 836	27	0.5	44145	87	AC010129	AC002102 Homo sapi
c 764	27	0.5	473	89	AF242445	L42376 Phytomonas	c 837	27	0.5	44145	87	AC010129	AC010129 Homo sapi
c 765	27	0.5	473	89	AF242445	AF242445 Homo sapi	c 838	27	0.5	44496	85	AC004760	AC010129 Homo sapi
c 766	27	0.5	578	91	F32281S09	AJ010984 Mus muscu	c 839	27	0.5	45186	91	AP001213	AC004760 Homo sapi
c 767	27	0.5	593	91	BC005936	AF232889 Homo sapi	c 840	27	0.5	45333	92	HS10618A	AP001213 Homo sapi
c 768	27	0.5	667	94	AF150090	BC005936 Homo sapi	c 841	27	0.5	46251	88	AD000671	AL049773 Human DNA
c 769	27	0.5	724	14	HVPSAG	AF150090 Mus muscu	c 842	27	0.5	48014	77	AC087054	AD000671 Homo sapi
c 770	27	0.5	730	97	HSU67222	X60158 H.vulgaris P	c 843	27	0.5	48716	77	AC087388	AC087054 Homo sapi
c 771	27	0.5	912	97	HSU67215	U67222 Human clone	c 844	27	0.5	51150	93	HSDJ913G4	AC087388 Homo sapi
c 772	27	0.5	1047	6	DD10DPG1	U67215 Human clone	c 845	27	0.5	51553	76	AC083779	AL049632 Human DNA
c 773	27	0.5	1055	94	BC002290	M30467 D.discoideu	c 846	27	0.5	52286	77	AC090001	AC083779 Homo sapi
c 774	27	0.5	1115	96	SCU96099	BC002290 Mus muscu	c 847	27	0.5	53038	78	AC090587	AC090001 Homo sapi
c 775	27	0.5	1136	9	A69662	U96099 Sarcophaga	c 848	27	0.5	54240	90	AL445215	AL045215 Human DNA
c 776	27	0.5	1144	85	AB045988	A69662 Sequence 15	c 849	27	0.5	56253	77	AC090174	AC090174 Homo sapi
c 777	27	0.5	1467	92	HSN223951	AB045988 Macaca fa	c 850	27	0.5	56835	69	AC025369	AC090174 Homo sapi
c 778	27	0.5	1481	94	AF029212	AJ223951 Homo sapi	c 851	27	0.5	57084	90	AL158813	AC025369 Homo sapi
c 779	27	0.5	1598	6	DDMFOSLC	AF029212 Rattus no	c 852	27	0.5	57304	85	AC004802	AL158813 Human DNA
c 780	27	0.5	1705	6	DDU20806	X54161 D.discoideu	c 853	27	0.5	58129	78	AC090404	AC004802 Homo sapi
c 781	27	0.5	1770	97	LEHBB	U20806 Dictyosteli	c 854	27	0.5	58129	78	AC090404	AC090404 Homo sapi
c 782	27	0.5	1835	13	AF281062	M15734 Lemur (brow	c 855	27	0.5	60742	90	AL445208	AL445208 Human DNA
c 783	27	0.5	2252	97	HUMCOL1A2	AF281062 Arabidops	c 856	27	0.5	61514	77	AC087697	AL445208 Human DNA
c 784	27	0.5	2251	88	AF068651	M55998 Human alpha	c 857	27	0.5	61544	77	AC090360	AC087697 Homo sapi
c 785	27	0.5	2454	97	HUMTSP	AF068651 Homo sapi	c 858	27	0.5	61795	78	AC090921	AC090360 Homo sapi
c 786	27	0.5	2618	14	CRCPRA	DI6517 Homo sapten	c 859	27	0.5	62232	63	AC015655	AC090921 Homo sapi
c 787	27	0.5	2785	94	AF162890S1	X69791 C.roseus cp	c 860	27	0.5	62275	76	AC079335	AC015655 Homo sapi
c 788	27	0.5	2848	94	BC004799	AF162890 Mus muscu	c 861	27	0.5	62366	71	AC036226	AC079335 Homo sapi
c 789	27	0.5	2917	94	AF225969	BC004799 Mus muscu	c 862	27	0.5	62964	76	AC080099	AC036226 Homo sapi
c 790	27	0.5	3447	9	A69663	AF225969 Mus muscu	c 863	27	0.5	63437	93	HSJ964N17	AC080099 Homo sapi
c 791	27	0.5	3450	8	XL004707	A69663 Sequence 16	c 864	27	0.5	63773	12	AC074025	AL050311 Human DNA
c 792	27	0.5	5759	97	HUMFBNIGNA	U04707 Xenopus lae	c 865	27	0.5	64058	96	AC084113	AC074025 Arabidops
c 793	27	0.5	6070	83	AF085746	L19896 Human fibri	c 866	27	0.5	64183	73	HSJ429G5	AC084113 Homo sapi
c 794	27	0.5	6728	93	HSPALICO	AF085746 Dario rer	c 867	27	0.5	64450	88	AC073893	AL078596 Human DNA
c 795	27	0.5	6913	97	HSU19765	274615 H.sapiens m	c 868	27	0.5	64814	14	NCB21321	AC073893 Homo sapi
c 796	27	0.5	10234	94	AF094611	U19765 Human nucle	c 869	27	0.5	64898	78	AC090735	AL355929 Neurospor
c 797	27	0.5	15302	90	AL163534	AF094611 Mus muscu	c 870	27	0.5	65034	64	AC016067	AC090735 Homo sapi
c 798	27	0.5	15405	89	AF220542	AL163534 Human DNA	c 871	27	0.5	65034	77	AC087290	AC016067 Homo sapi
c 799	27	0.5	16549	15	T31J12	AF220542 Homo sapi	c 872	27	0.5	65034	77	AC087290	AC087290 Homo sapi
c 800	27	0.5	17906	94	AF071560	AC006416 Arabidops	c 873	27	0.5	65427	78	AC091274	AC087290 Homo sapi
c 801	27	0.5	18609	88	AF071560	AF071560 Mus muscu	c 874	27	0.5	65472	78	AC091274	AC091274 Mus muscu
c 802	27	0.5	25483	91	AL450431	AF071560 Homo sapi	c 875	27	0.5	65663	78	AC091147	AC091274 Homo sapi
c 803	27	0.5	28047	92	HS44083	AL450431 Human DNA	c 876	27	0.5	66227	70	AC026534	AC091147 Homo sapi
c 804	27	0.5	29122	91	AL513125	AL022331 Human DNA	c 877	27	0.5	66585	91	AL512378	AC026534 Homo sapi
c 805	27	0.5	30758	89	AF278704	AL513125 Human DNA	c 878	27	0.5	66975	89	AL135791	AL512378 Human DNA
c 806	27	0.5	32239	94	AB034241	AF278704 Homo sapi	c 879	27	0.5	67150	69	AC025606	AL135791 Homo sapi
c 807	27	0.5	34164	78	AC090913	AB034241 Rattus no	c 880	27	0.5	67722	77	AC087747	AC025606 Homo sapi
c 808	27	0.5	34544	6	CELY5868A	AC090913 Homo sapi	c 881	27	0.5	68037	66	AC021757	AC087747 Homo sapi
c 809	27	0.5	35687	85	AC004779	AC06808 Caenorhab	c 882	27	0.5	68279	77	AC090165	AC021757 Homo sapi
c 810	27	0.5	35882	97	HSU41384	AC004779 Homo sapi	c 883	27	0.5	68407	72	AC048372	AC090165 Homo sapi
c 811	27	0.5	36255	64	AC016208	U41384 Human small	c 884	27	0.5	68763	78	AC090718	AC048372 Homo sapi
						AC016208 Homo sapi							

593	28	0.5	179473	64	AC016310	AC016310 Homo sapi	666	28	0.5	190709	60	AC008435	AC008435 Homo sapi
594	28	0.5	179485	86	AC005939	AC005939 Homo sapi	c 667	28	0.5	190821	86	AC006111	AC006111 Homo sapi
595	28	0.5	179597	73	AC068447	AC068447 Homo sapi	c 668	28	0.5	190907	85	AC005261	AC005261 Homo sapi
596	28	0.5	179613	81	AL390881	AL390881 Homo sapi	c 669	28	0.5	191422	67	AC022318	AC022318 Homo sapi
597	28	0.5	179704	70	AC026627	AC026627 Homo sapi	c 670	28	0.5	191877	86	AC007207	AC007207 Homo sapi
598	28	0.5	179797	75	AC073865	AC073865 Homo sapi	671	28	0.5	191904	86	AC005699	AC005699 Homo sapi
599	28	0.5	179929	86	AC006270	AC006270 Homo sapi	c 672	28	0.5	192126	91	DJ270M14	AF107885 Homo sapi
600	28	0.5	179955	87	AC023247	AC023247 Homo sapi	c 673	28	0.5	192218	75	AC074021	AC074021 Homo sapi
601	28	0.5	180038	73	AC068393	AC068393 Homo sapi	674	28	0.5	192318	95	AL355341	AL355341 Homo sapi
602	28	0.5	180129	75	AC074173	AC074173 Mus muscu	675	28	0.5	192610	79	AL355341	AL355341 Homo sapi
603	28	0.5	180254	76	AC084001	AC084001 Homo sapi	676	28	0.5	192782	75	AC079276	AC079276 Mus muscu
604	28	0.5	180397	86	AC007035	AC007035 Homo sapi	677	28	0.5	193020	89	AL138759	AL138759 Human DNA
605	28	0.5	180435	72	AC060812	AC060812 Homo sapi	678	28	0.5	193086	78	AF286887	AF286887 Homo sapi
606	28	0.5	180596	87	AC009570	AC009570 Homo sapi	c 679	28	0.5	193513	75	AC074372	AC074372 Homo sapi
607	28	0.5	180686	94	AF267747	AF267747 Mus muscu	c 680	28	0.5	193573	78	AL138921	AL138921 Homo sapi
608	28	0.5	180791	68	AC023977	AC023977 Homo sapi	c 681	28	0.5	194032	65	AC019150	AC019150 Homo sapi
609	28	0.5	180907	72	AC048383	AC048383 Homo sapi	c 682	28	0.5	194036	81	AL513479	AL513479 Homo sapi
610	28	0.5	181485	68	AC024036	AC024036 Homo sapi	c 683	28	0.5	194497	88	AC021887	AC021887 Homo sapi
611	28	0.5	181487	78	AL139410	AL139410 Homo sapi	c 684	28	0.5	194612	90	AL357354	AL357354 Human DNA
612	28	0.5	181823	71	AC040167	AC040167 Homo sapi	c 685	28	0.5	194794	62	AC012371	AC012371 Homo sapi
613	28	0.5	181930	78	AC091297	AC091297 Sus scrofa	686	28	0.5	194935	69	AC026113	AC026113 Homo sapi
614	28	0.5	182158	81	AL391116	AL391116 Homo sapi	687	28	0.5	194942	77	AC087884	AC087884 Homo sapi
615	28	0.5	182189	86	AC007333	AC007333 Homo sapi	688	28	0.5	195470	76	AC079367	AC079367 Mus muscu
616	28	0.5	182262	68	AC023963	AC023963 Homo sapi	689	28	0.5	195529	69	AC025612	AC025612 Homo sapi
617	28	0.5	182303	60	AC007638	AC007638 Homo sapi	c 690	28	0.5	196379	78	AL139319	AL139319 Homo sapi
618	28	0.5	182799	74	AC073526	AC073526 Homo sapi	c 691	28	0.5	196751	69	AC025494	AC025494 Homo sapi
619	28	0.5	182919	73	AC068779	AC068779 Homo sapi	c 692	28	0.5	196986	82	AF000941	AF000941 Homo sapi
620	28	0.5	183029	67	AC022067	AC022067 Homo sapi	c 693	28	0.5	197054	82	AL590036	AL590036 Homo sapi
621	28	0.5	183490	71	AC034244	AC034244 Homo sapi	c 694	28	0.5	197189	65	AC019168	AC019168 Homo sapi
622	28	0.5	183550	64	AC015913	AC015913 Homo sapi	c 695	28	0.5	197523	65	AC018883	AC018883 Homo sapi
623	28	0.5	183680	60	AC009077	AC009077 Homo sapi	c 696	28	0.5	197692	61	AC009554	AC009554 Homo sapi
624	28	0.5	183692	81	AL451125	AL451125 Homo sapi	697	28	0.5	198333	81	AL445673	AL445673 Homo sapi
625	28	0.5	183916	89	AF111169	AF111169 Homo sapi	c 698	28	0.5	198388	74	AC073426	AC073426 Homo sapi
626	28	0.5	184012	87	AC017080	AC017080 Homo sapi	c 699	28	0.5	199243	71	AC040925	AC040925 Homo sapi
627	28	0.5	184137	70	AC026472	AC026472 Homo sapi	c 700	28	0.5	200490	69	AC025548	AC025548 Homo sapi
628	28	0.5	184148	77	AC084413	AC084413 Homo sapi	c 701	28	0.5	200688	74	AC073266	AC073266 Homo sapi
629	28	0.5	184181	90	AL158141	AL158141 Human DNA	702	28	0.5	200891	91	AL451146	AL451146 Human DNA
630	28	0.5	184655	76	AC079379	AC079379 Homo sapi	703	28	0.5	201143	92	HS150C2	AL451146 Human DNA
631	28	0.5	184834	88	AC026122	AC026122 Homo sapi	c 704	28	0.5	201290	77	AC084375	AC084375 Homo sapi
632	28	0.5	184840	66	AC021422	AC021422 Homo sapi	c 705	28	0.5	201685	79	AL353743	AL353743 Homo sapi
633	28	0.5	184852	79	AL162732	AL162732 Homo sapi	c 706	28	0.5	202267	91	DJ293M10	AF111167 Homo sapi
634	28	0.5	184864	63	AC013553	AC013553 Homo sapi	c 707	28	0.5	202596	65	AC018441	AC018441 Homo sapi
635	28	0.5	185019	66	AC021231	AC021231 Homo sapi	c 708	28	0.5	202593	72	AC063955	AC063955 Homo sapi
636	28	0.5	185143	69	AC025861	AC025861 Homo sapi	c 709	28	0.5	202719	86	AC006379	AC006379 Homo sapi
637	28	0.5	185173	60	AC009056	AC009056 Homo sapi	c 710	28	0.5	202986	72	AC062014	AC062014 Homo sapi
638	28	0.5	185186	79	AL158826	AL158826 Homo sapi	c 711	28	0.5	203046	60	AC006342	AC006342 Homo sapi
639	28	0.5	185203	77	AL161649	AL161649 Homo sapi	712	28	0.5	203274	80	AL356957	AL356957 Homo sapi
640	28	0.5	185207	79	AC084350	AC084350 Homo sapi	c 713	28	0.5	204201	81	AL589666	AL589666 Homo sapi
641	28	0.5	185331	62	AC012059	AC012059 Homo sapi	c 714	28	0.5	204267	65	AC019071	AC019071 Homo sapi
642	28	0.5	185666	72	AC048389	AC048389 Homo sapi	c 715	28	0.5	204809	66	AC021133	AC021133 Homo sapi
643	28	0.5	185688	65	AC018805	AC018805 Homo sapi	c 716	28	0.5	205307	61	AC009968	AC009968 Homo sapi
644	28	0.5	185764	69	AC025289	AC025289 Homo sapi	c 717	28	0.5	205430	83	AP002756	AP002756 Homo sapi
645	28	0.5	185820	91	HS1009E24	AL109804 Human DNA	718	28	0.5	205499	68	AC023352	AC023352 Homo sapi
646	28	0.5	185960	60	AC008245	AC008245 Homo sapi	719	28	0.5	205993	62	AC011952	AC011952 Homo sapi
647	28	0.5	185994	60	AC002042	AC002042 Homo sapi	c 720	28	0.5	206578	70	AC026529	AC026529 Homo sapi
648	28	0.5	186221	62	AC012530	AC012530 Homo sapi	c 721	28	0.5	206647	91	AP002898	AP002898 Homo sapi
649	28	0.5	186321	86	AC006511	AC006511 Homo sapi	c 722	28	0.5	206848	72	AC063965	AC063965 Homo sapi
650	28	0.5	187386	69	AC025554	AC025554 Homo sapi	c 723	28	0.5	207170	83	AP003484	AP003484 Homo sapi
651	28	0.5	187451	74	AC073216	AC073216 Homo sapi	c 724	28	0.5	208145	72	AC053504	AC053504 Homo sapi
652	28	0.5	187647	74	AC073220	AC073220 Homo sapi	c 725	28	0.5	208763	79	AL353791	AL353791 Homo sapi
653	28	0.5	187877	67	AC023251	AC023251 Homo sapi	c 726	28	0.5	209303	66	AC020690	AC020690 Homo sapi
654	28	0.5	187955	73	AC067759	AC067759 Homo sapi	c 727	28	0.5	209845	70	AC026401	AC026401 Homo sapi
655	28	0.5	188152	69	AC025847	AC025847 Homo sapi	c 728	28	0.5	210946	79	AL353620	AL353620 Homo sapi
656	28	0.5	188603	76	AC079790	AC079790 Homo sapi	729	28	0.5	211442	64	AC016886	AC016886 Homo sapi
657	28	0.5	188684	90	AL355312	AL355312 Human DNA	730	28	0.5	216288	80	AL390116	AL390116 Homo sapi
658	28	0.5	189067	64	AC016810	AC016810 Homo sapi	731	28	0.5	216387	91	DJ534K4	AF109907 Homo sapi
659	28	0.5	189287	73	AC068362	AC068362 Homo sapi	c 732	28	0.5	216842	63	AC013726	AC013726 Homo sapi
660	28	0.5	189592	65	AC019146	AC019146 Homo sapi	c 733	28	0.5	218836	87	AC016734	AC016734 Homo sapi
661	28	0.5	189730	76	AC084044	AC084044 Mus muscu	c 734	28	0.5	219139	77	AC008418	AC008418 Homo sapi
662	28	0.5	189835	66	AC021011	AC021011 Homo sapi	735	28	0.5	219820	87	AC008738	AC008738 Homo sapi
663	28	0.5	189953	90	AL445435	AL445435 Human DNA	c 736	28	0.5	219935	85	AC005015	AC005015 Homo sapi
664	28	0.5	190000	60	AC004479	AC004479 Homo sapi	737	28	0.5	220553	82	AP001025	AP001025 Homo sapi
665	28	0.5	190440	90	AL357515	AL357515 Human DNA	c 738	28	0.5	221484	65	AC019095	AC019095 Homo sapi

C 447	28	0.5	158666	67	AC022394	AC022394	Homo sapi	520	28	0.5	171216	79	AL354669	AL354669	Homo sapi
C 448	28	0.5	158760	69	AC026098	AC026098	Homo sapi	C 521	28	0.5	171292	67	AC021933	AC021933	Homo sapi
C 449	28	0.5	159096	90	AL161781	AL161781	Human DNA	C 522	28	0.5	171300	61	AC010941	AC010941	Homo sapi
C 450	28	0.5	159175	74	AC069421	AC069421	Homo sapi	C 523	28	0.5	171708	88	AC067956	AC067956	Homo sapi
C 451	28	0.5	159508	65	AC018573	AC018573	Homo sapi	C 524	28	0.5	171732	80	AL356152	AL356152	Homo sapi
C 452	28	0.5	159705	79	AL162613	AL162613	Homo sapi	C 525	28	0.5	171732	67	AC021994	AC021994	Homo sapi
C 453	28	0.5	159859	75	AC074010	AC074010	Homo sapi	C 526	28	0.5	171767	77	AC089988	AC089988	Homo sapi
C 454	28	0.5	160109	90	AL161727	AL161727	Human DNA	C 527	28	0.5	171805	87	AC020716	AC020716	Homo sapi
C 455	28	0.5	160169	64	AC016660	AC016660	Homo sapi	C 528	28	0.5	171902	82	AP000846	AP000846	Homo sapi
C 456	28	0.5	160193	69	AC025605	AC025605	Homo sapi	C 529	28	0.5	171908	69	AC025702	AC025702	Homo sapi
C 457	28	0.5	160383	63	AC015523	AC015523	Homo sapi	C 530	28	0.5	172111	65	AC018498	AC018498	Homo sapi
C 458	28	0.5	160590	61	AC009663	AC009663	Homo sapi	C 531	28	0.5	172144	73	AC067835	AC067835	Homo sapi
C 459	28	0.5	160644	88	AC068770	AC068770	Homo sapi	C 532	28	0.5	172219	81	AL442004	AL442004	Homo sapi
C 460	28	0.5	160696	88	AC090937	AC090937	Homo sapi	C 533	28	0.5	172291	86	AC006337	AC006337	Homo sapi
C 461	28	0.5	160722	64	AC016963	AC016963	Homo sapi	C 534	28	0.5	172403	64	AC017061	AC017061	Homo sapi
C 462	28	0.5	160759	83	AP002376	AP002376	Homo sapi	C 535	28	0.5	172472	60	AC008652	AC008652	Homo sapi
C 463	28	0.5	160831	69	AC025015	AC025015	Homo sapi	C 536	28	0.5	172526	70	AC026934	AC026934	Homo sapi
C 464	28	0.5	160892	81	AL512410	AL512410	Homo sapi	C 537	28	0.5	172620	80	AL359472	AL359472	Homo sapi
C 465	28	0.5	160956	63	AC013637	AC013637	Homo sapi	C 538	28	0.5	172751	65	AC018889	AC018889	Homo sapi
C 466	28	0.5	161155	88	AF017104	AF017104	Homo sapi	C 539	28	0.5	172813	63	AC015720	AC015720	Homo sapi
C 467	28	0.5	161356	76	AC079848	AC079848	Homo sapi	C 540	28	0.5	173040	87	AC012442	AC012442	Homo sapi
C 468	28	0.5	161514	88	AC078845	AC078845	Homo sapi	C 541	28	0.5	173081	78	AL137001	AL137001	Homo sapi
C 469	28	0.5	161810	60	AC009082	AC009082	Homo sapi	C 542	28	0.5	173390	91	CNS05TCM	CNS05TCM	Human chr
C 470	28	0.5	161979	81	AL451081	AL451081	Homo sapi	C 543	28	0.5	173427	71	AC032044	AC032044	Homo sapi
C 471	28	0.5	162075	92	HS127D3	HS127D3	Homo sapi	C 544	28	0.5	173853	65	AC018999	AC018999	Homo sapi
C 472	28	0.5	162081	86	AC007542	AC007542	Homo sapi	C 545	28	0.5	174013	81	AL451062	AL451062	Homo sapi
C 473	28	0.5	162105	70	AC026483	AC026483	Homo sapi	C 546	28	0.5	174250	72	AC063950	AC063950	Homo sapi
C 474	28	0.5	162416	88	AC026882	AC026882	Homo sapi	C 547	28	0.5	174543	61	AC009127	AC009127	Homo sapi
C 475	28	0.5	162426	82	AL590377	AL590377	Homo sapi	C 548	28	0.5	174799	83	AP003124	AP003124	Homo sapi
C 476	28	0.5	162602	74	AC073633	AC073633	Homo sapi	C 549	28	0.5	174823	64	AC016106	AC016106	Homo sapi
C 477	28	0.5	162617	86	AC007501	AC007501	Homo sapi	C 550	28	0.5	174832	79	AL353718	AL353718	Homo sapi
C 478	28	0.5	162797	67	AC023219	AC023219	Homo sapi	C 551	28	0.5	174869	82	AP001840	AP001840	Homo sapi
C 479	28	0.5	162829	81	AL513185	AL513185	Homo sapi	C 552	28	0.5	174933	87	AC009087	AC009087	Homo sapi
C 480	28	0.5	163567	80	AL365325	AL365325	Homo sapi	C 553	28	0.5	175093	82	AP001363	AP001363	Homo sapi
C 481	28	0.5	163608	69	AC026081	AC026081	Homo sapi	C 554	28	0.5	175109	79	AL354659	AL354659	Homo sapi
C 482	28	0.5	164039	74	AC069354	AC069354	Homo sapi	C 555	28	0.5	175129	66	AC021490	AC021490	Homo sapi
C 483	28	0.5	164247	79	AL158082	AL158082	Homo sapi	C 556	28	0.5	175363	76	AC079897	AC079897	Homo sapi
C 484	28	0.5	164347	71	AC034124	AC034124	Homo sapi	C 557	28	0.5	175611	66	AC020621	AC020621	Homo sapi
C 485	28	0.5	164352	68	AC024008	AC024008	Homo sapi	C 558	28	0.5	175706	89	AL135783	AL135783	Human DNA
C 486	28	0.5	164463	80	AL365210	AL365210	Homo sapi	C 559	28	0.5	176353	65	AC019144	AC019144	Homo sapi
C 487	28	0.5	164494	87	AC020898	AC020898	Homo sapi	C 560	28	0.5	176388	69	AC025273	AC025273	Homo sapi
C 488	28	0.5	164757	81	AL391215	AL391215	Homo sapi	C 561	28	0.5	176496	69	AC025517	AC025517	Homo sapi
C 489	28	0.5	164823	87	AC009567	AC009567	Homo sapi	C 562	28	0.5	176630	79	AL353600	AL353600	Homo sapi
C 490	28	0.5	165255	66	AC021421	AC021421	Homo sapi	C 563	28	0.5	176638	61	AC010809	AC010809	Homo sapi
C 491	28	0.5	165287	77	AC087742	AC087742	Homo sapi	C 564	28	0.5	176697	81	AL391561	AL391561	Homo sapi
C 492	28	0.5	165390	78	AC090529	AC090529	Homo sapi	C 565	28	0.5	176705	63	AC013715	AC013715	Homo sapi
C 493	28	0.5	165537	73	AC068884	AC068884	Homo sapi	C 566	28	0.5	176773	61	AC009867	AC009867	Homo sapi
C 494	28	0.5	166154	69	AC025779	AC025779	Homo sapi	C 567	28	0.5	176787	78	AL138891	AL138891	Homo sapi
C 495	28	0.5	166549	69	AC025990	AC025990	Homo sapi	C 568	28	0.5	176833	91	CNS00M8V	CNS00M8V	Human chr
C 496	28	0.5	166582	77	AC084398	AC084398	Homo sapi	C 569	28	0.5	176929	85	AC004887	AC004887	Homo sapi
C 497	28	0.5	166797	80	AL359094	AL359094	Homo sapi	C 570	28	0.5	177080	73	AC067925	AC067925	Homo sapi
C 498	28	0.5	167327	80	AL355979	AL355979	Homo sapi	C 571	28	0.5	177120	61	AC009967	AC009967	Homo sapi
C 499	28	0.5	167603	73	AC068261	AC068261	Homo sapi	C 572	28	0.5	177198	72	AC055732	AC055732	Homo sapi
C 500	28	0.5	167712	68	AC023809	AC023809	Mus muscu	C 573	28	0.5	177341	22	AC012036	AC012036	Homo sapi
C 501	28	0.5	168111	69	AC025788	AC025788	Homo sapi	C 574	28	0.5	177364	86	AC005409	AC005409	Homo sapi
C 502	28	0.5	168210	87	AC018719	AC018719	Homo sapi	C 575	28	0.5	177393	75	AC079019	AC079019	Homo sapi
C 503	28	0.5	168347	69	AC025336	AC025336	Homo sapi	C 576	28	0.5	177403	73	AC067739	AC067739	Homo sapi
C 504	28	0.5	168621	72	AC046173	AC046173	Homo sapi	C 577	28	0.5	177428	94	AC026384	AC026384	Mus muscu
C 505	28	0.5	168665	72	AC053491	AC053491	Homo sapi	C 578	28	0.5	177449	62	AC012322	AC012322	Homo sapi
C 506	28	0.5	168749	60	AC008573	AC008573	Homo sapi	C 579	28	0.5	177585	86	AC005826	AC005826	Homo sapi
C 507	28	0.5	169162	62	AC012479	AC012479	Homo sapi	C 580	28	0.5	177629	72	AC060829	AC060829	Homo sapi
C 508	28	0.5	169504	91	CNS010DTV	CNS010DTV	Human chr	C 581	28	0.5	177784	62	AC011737	AC011737	Homo sapi
C 509	28	0.5	169524	67	AC023111	AC023111	Homo sapi	C 582	28	0.5	177843	68	AC024446	AC024446	Homo sapi
C 510	28	0.5	169622	80	AL359761	AL359761	Homo sapi	C 583	28	0.5	178183	93	HS342D11	HS342D11	Human DNA
C 511	28	0.5	170278	66	AC020586	AC020586	Homo sapi	C 584	28	0.5	178414	69	AC025341	AC025341	Homo sapi
C 512	28	0.5	170278	66	AC020586	AC020586	Homo sapi	C 585	28	0.5	178874	70	AC027085	AC027085	Homo sapi
C 513	28	0.5	170399	93	HSJ7680D5	HSJ7680D5	Human DNA	C 586	28	0.5	178903	64	AC016453	AC016453	Homo sapi
C 514	28	0.5	170420	62	AC011612	AC011612	Homo sapi	C 587	28	0.5	179064	90	AL353807	AL353807	Human DNA
C 515	28	0.5	170609	77	AC084863	AC084863	Papio cyn	C 588	28	0.5	179151	79	AL353578	AL353578	Homo sapi
C 516	28	0.5	170623	90	AL391375	AL391375	Human DNA	C 589	28	0.5	179167	73	AC068891	AC068891	Homo sapi
C 517	28	0.5	170923	64	AC016181	AC016181	Homo sapi	C 590	28	0.5	179245	97	HUAC002492	HUAC002492	Human chr
C 518	28	0.5	171115	65	AC018846	AC018846	Homo sapi	C 591	28	0.5	179318	84	CNS05TDF	CNS05TDF	Homo sapi
C 519	28	0.5	171135	81	AL441886	AL441886	Homo sapi	C 592	28	0.5	179451	83	AP003476	AP003476	Homo sapi

C 301	28	0.5	66671	78	AC090795	Homo sapi	AC090795	0.5	138041	91	CNS01DU1	AL133245	BAC seque
C 302	28	0.5	70184	87	AC012441	Homo sapi	AC012441	0.5	138532	85	AC004812	Homo sapi	AC004812
C 303	28	0.5	72254	90	AL391646	Human DNA	AL391646	0.5	138613	77	AC084702	Homo sapi	AC084702
C 304	28	0.5	72807	68	AC023525	Homo sapi	AC023525	0.5	138814	73	AC068997	Homo sapi	AC068997
C 305	28	0.5	73403	79	AL355505	Homo sapi	AL355505	0.5	138839	97	HUAC002400	Human Chr	AC002400
C 306	28	0.5	74961	79	AL162258	Homo sapi	AL162258	0.5	139485	69	AC025597	Homo sapi	AC025597
C 307	28	0.5	75110	74	AC069115	Homo sapi	AC069115	0.5	139627	90	AL357117	Human DNA	AL357117
C 308	28	0.5	76241	93	HSJ336M4	Human DNA	AC069175	0.5	139904	92	HS97D16	Homo sapi	AL009179
C 309	28	0.5	76660	69	AC025153	Homo sapi	AC025153	0.5	139973	67	AC022650	Homo sapi	AC022650
C 310	28	0.5	79612	85	AC004073	Human Chr	AC004073	0.5	140738	89	AF129408	Homo sapi	AF129408
C 311	28	0.5	81854	86	AC006396	Homo sapi	AC006396	0.5	141277	90	AL357045	Human DNA	AL357045
C 312	28	0.5	82001	12	AB005242	Arabidops	AB005242	0.5	141514	80	AL359095	Homo sapi	AL359095
C 313	28	0.5	82404	89	AL133399	Human DNA	AL133399	0.5	141614	82	AL590244	Homo sapi	AL590244
C 314	28	0.5	84122	89	AL133463	Human DNA	AL133463	0.5	142908	81	AL513172	Homo sapi	AL513172
C 315	28	0.5	86064	81	AL391554	Homo sapi	AL391554	0.5	143183	70	AC026927	Homo sapi	AC026927
C 316	28	0.5	86519	91	AP001860	Homo sapi	AP001860	0.5	143200	88	AC022081	Homo sapi	AC022081
C 317	28	0.5	87039	12	AC016529	Homo sapi	AC016529	0.5	143369	86	AC006316	Homo sapi	AC006316
C 318	28	0.5	89820	86	AC007868	Genomic S	AC007868	0.5	143754	80	AL359093	Homo sapi	AL359093
C 319	28	0.5	92969	93	HSJ687F11	Human DNA	AL117334	0.5	143754	80	AL359093	Homo sapi	AL359093
C 320	28	0.5	95565	85	AC000379	Human Chr	AC000379	0.5	144114	79	AL161453	Homo sapi	AL161453
C 321	28	0.5	96202	89	AL138824	Human DNA	AL138824	0.5	144168	66	AC021203	Homo sapi	AC021203
C 322	28	0.5	96726	84	HS1178F13	Homo sapi	AL118500	0.5	144660	60	AC008650	Homo sapi	AC008650
C 323	28	0.5	97982	91	AP000687	Homo sapi	AP000687	0.5	144967	60	AC008963	Homo sapi	AC008963
C 324	28	0.5	98869	62	AC011496	Homo sapi	AC011496	0.5	145253	86	AC006329	Homo sapi	AC006329
C 325	28	0.5	98933	88	AC025435	Homo sapi	AC025435	0.5	146093	78	AL139106	Homo sapi	AL139106
C 326	28	0.5	99011	60	AC004937	Homo sapi	AC004937	0.5	146170	75	AC079224	Homo sapi	AC079224
C 327	28	0.5	100301	92	HS317E23	Human DNA	AL020996	0.5	146260	60	AC008769	Homo sapi	AC008769
C 328	28	0.5	100506	83	AP002020	Homo sapi	AP002020	0.5	146372	90	AL359174	Human DNA	AL359174
C 329	28	0.5	101370	85	AC003962	Homo sapi	AC003962	0.5	146385	87	AC010534	Homo sapi	AC010534
C 330	28	0.5	102910	85	AB041512	Homo sapi	AB041512	0.5	147094	62	AC012027	Homo sapi	AC012027
C 331	28	0.5	104025	89	AL138679	Human DNA	AL138679	0.5	147786	61	AC010299	Homo sapi	AC010299
C 332	28	0.5	104147	85	AC004703	Homo sapi	AC004703	0.5	147945	62	AC012244	Homo sapi	AC012244
C 333	28	0.5	104269	70	AC026448	Homo sapi	AC026448	0.5	148515	83	CNS01RGK	Homo sapi	AL158059
C 334	28	0.5	106830	63	AC013394	Homo sapi	AC013394	0.5	148641	92	HS1119D9	Human DNA	AL031652
C 335	28	0.5	107260	90	AL139352	Human DNA	AL139352	0.5	149087	90	AL157826	Human DNA	AL157826
C 336	28	0.5	107418	92	HS551E13	Human DNA	AL022163	0.5	149253	62	AC012245	Homo sapi	AC012245
C 337	28	0.5	107429	93	HSJ214M20	Human DNA	AL121969	0.5	149411	65	AC019029	Homo sapi	AC019029
C 338	28	0.5	108090	87	AC009533	Homo sapi	AC009533	0.5	149480	97	HU095740	Human Chrom	U95740
C 339	28	0.5	108400	85	AC005193	Homo sapi	AC005193	0.5	149605	86	AC005522	Homo sapi	AC005522
C 340	28	0.5	109524	93	HSJ305B16	Human DNA	AL047972	0.5	150765	68	AC024326	Homo sapi	AC024326
C 341	28	0.5	110000	62	AC011600.2	Continuation (3 of	Continuation (3 of	0.5	150846	89	AL138724	Human DNA	AL138724
C 342	28	0.5	110000	64	AC015817.0	Homo sapi	AC015817	0.5	151851	84	CNS01RHL	Homo sapi	AL161804
C 343	28	0.5	110000	85	AC005079.0	Homo sapi	AC005079	0.5	151933	74	AC073199	Homo sapi	AC073199
C 344	28	0.5	110384	85	AC000117	Homo sapi	AC000117	0.5	152058	69	AC025232	Homo sapi	AC025232
C 345	28	0.5	111035	88	AC022405	Homo sapi	AC022405	0.5	152514	66	AC021329	Homo sapi	AC021329
C 346	28	0.5	111868	64	AC015848	Homo sapi	AC015848	0.5	152612	72	AC053515	Homo sapi	AC053515
C 347	28	0.5	111977	82	AP000717	Homo sapi	AP000717	0.5	152798	85	AC004057	Homo sapi	AC004057
C 348	28	0.5	112705	83	AP002025	Homo sapi	AP002025	0.5	153225	63	AC013581	Homo sapi	AC013581
C 349	28	0.5	112769	93	HSJ635E18	Human DNA	AL109811	0.5	153493	69	AC026215	Homo sapi	AC026215
C 350	28	0.5	114317	22	AC012017	Homo sapi	AC012017	0.5	154788	92	HS235F22	Human DNA	AL024508
C 351	28	0.5	115641	15	U95973	Arabidopsis	U95973	0.5	154848	63	AC012616	Homo sapi	AC012616
C 352	28	0.5	124474	90	AL158012	Human DNA	AL158012	0.5	155112	80	AL357521	Homo sapi	AL357521
C 353	28	0.5	125202	85	AC004858	Homo sapi	AC004858	0.5	155493	87	AC008825	Homo sapi	AC008825
C 354	28	0.5	127178	85	AC005160	Homo sapi	AC005160	0.5	155745	67	AC021941	Homo sapi	AC021941
C 355	28	0.5	127606	80	AL356298	Homo sapi	AL356298	0.5	155745	67	AC021941	Homo sapi	AC021941
C 356	28	0.5	127867	92	HS694B14	Human DNA	AL031673	0.5	155868	79	AL353779	Homo sapi	AL353779
C 357	28	0.5	129491	78	AL138842	Homo sapi	AL138842	0.5	156159	67	AC023325	Homo sapi	AC023325
C 358	28	0.5	131159	82	AP000743	Homo sapi	AP000743	0.5	156270	76	AC083951	Homo sapi	AC083951
C 359	28	0.5	131245	89	AL133320	Human DNA	AL133320	0.5	156372	78	AL139138	Homo sapi	AL139138
C 360	28	0.5	131587	80	AL358075	Homo sapi	AL358075	0.5	157248	85	AC002312	Human Chr	AC002312
C 361	28	0.5	132275	72	AC048388	Homo sapi	AC048388	0.5	157302	81	AL445209	Homo sapi	AL445209
C 362	28	0.5	132353	60	AC002355	Homo sapi	AC002355	0.5	157419	69	AC025389	Homo sapi	AC025389
C 363	28	0.5	132805	92	HS339A18	Human DNA s	297054	0.5	157666	86	AC006989	Homo sapi	AC006989
C 364	28	0.5	135734	60	AC002346	Homo sapi	AC002346	0.5	157814	71	AC037460	Homo sapi	AC037460
C 365	28	0.5	135805	74	AC069511	Homo sapi	AC069511	0.5	157871	87	AC012099	Homo sapi	AC012099
C 366	28	0.5	135961	66	AC021288	Homo sapi	AC021288	0.5	157930	81	AL391065	Homo sapi	AL391065
C 367	28	0.5	136124	92	HS394P21	Homo sapi	AL021528	0.5	157949	90	AL355578	Human DNA	AL355578
C 368	28	0.5	136385	86	AC006006	Homo sapi	AC006006	0.5	158116	79	AL353627	Homo sapi	AL353627
C 369	28	0.5	136462	87	AC010142	Homo sapi	AC010142	0.5	158207	67	AC021942	Homo sapi	AC021942
C 370	28	0.5	137557	87	AC009414	Homo sapi	AC009414	0.5	158276	71	AC027709	Homo sapi	AC027709
C 371	28	0.5	137604	67	AC022275	Homo sapi	AC022275	0.5	158297	81	AL513545	Homo sapi	AL513545
C 372	28	0.5	137658	92	HS1022J11	Human DNA	AL049765	0.5	158608	70	AC026834	Homo sapi	AC026834
C 373	28	0.5	137940	63	AC015612	Homo sapi	AC015612	0.5	158620	74	AC069522	Homo sapi	AC069522

155	29	0.6	159487	71	AC036214	Homo sapi	c 228	29	0.6	199988	68	AC023545	Homo sapi
156	29	0.6	160228	81	AL512372	Homo sapi	c 229	29	0.6	200242	81	AC010895	Homo sapi
157	29	0.6	161514	88	AC078845	Homo sapi	c 230	29	0.6	200594	86	AC005632	Homo sapi
158	29	0.6	161940	87	AC078838	Homo sapi	c 231	29	0.6	202004	97	HUAF001549	Human Chr
159	29	0.6	162339	75	AC022960	Homo sapi	c 232	29	0.6	202092	80	AL558784	Homo sapi
160	29	0.6	162470	86	AC007664	Homo sapi	c 233	29	0.6	202121	81	AL512490	Homo sapi
161	29	0.6	162922	62	AC011281	Homo sapi	c 234	29	0.6	202380	82	AL589902	Mus muscu
162	29	0.6	163282	64	AC016438	Homo sapi	c 235	29	0.6	203790	87	AC010422	Homo sapi
163	29	0.6	165872	69	AC025089	Homo sapi	c 236	29	0.6	204122	81	AL450123	Homo sapi
164	29	0.6	166297	62	AC011271	Homo sapi	c 237	29	0.6	204224	77	AC087891	Mus muscu
165	29	0.6	166884	70	AC027147	Homo sapi	c 238	29	0.6	205778	61	AC009624	Homo sapi
166	29	0.6	166884	70	AC027147	Homo sapi	c 239	29	0.6	207641	63	AC015800	Mus muscu
167	29	0.6	167460	97	UB2695	Homo sapien	c 240	29	0.6	207653	79	AL355614	Homo sapi
168	29	0.6	167629	69	AC024974	Homo sapi	c 241	29	0.6	207693	82	AL590556	Homo sapi
169	29	0.6	168043	62	AC011935	Homo sapi	c 242	29	0.6	208006	77	AC087877	Mus muscu
170	29	0.6	169677	71	AC036215	Homo sapi	c 243	29	0.6	208632	74	AC073711	Mus muscu
171	29	0.6	169819	72	AC035307	Homo sapi	c 244	29	0.6	209759	82	AP001029	Homo sapi
172	29	0.6	170128	60	AC005059	Homo sapi	c 245	29	0.6	210164	75	AC074314	Mus muscu
173	29	0.6	170197	61	AC010545	Homo sapi	c 246	29	0.6	215734	75	AC073828	Mus muscu
174	29	0.6	171049	82	AL590022	Homo sapi	c 247	29	0.6	219447	97	HUMFLNG6PD	
175	29	0.6	171140	88	AC069026	Homo sapi	c 248	29	0.6	220170	74	AC073511	Mus muscu
176	29	0.6	171490	67	AC022481	Homo sapi	c 249	29	0.6	227525	81	AL445524	Homo sapi
177	29	0.6	171549	62	AC011846	Homo sapi	c 250	29	0.6	230037	79	AL159174	Homo sapi
178	29	0.6	172241	86	AC008154	Homo sapi	c 251	29	0.6	231669	73	AC068067	Mus muscu
179	29	0.6	172332	61	AC009122	Homo sapi	c 252	29	0.6	249487	94	MMU251788	Mus muscu
180	29	0.6	173804	80	AL357835	Homo sapi	c 253	29	0.6	249487	94	MMU251835	Mus muscu
181	29	0.6	174024	66	AC021143	Homo sapi	c 254	29	0.6	262415	67	AC023143	Homo sapi
182	29	0.6	174034	63	AC013438	Homo sapi	c 255	29	0.6	265016	60	AC008383	Homo sapi
183	29	0.6	174409	61	AC010788	Homo sapi	c 256	29	0.6	269576	74	AC073684	Mus muscu
184	29	0.6	175986	67	AC023071	Homo sapi	c 257	29	0.6	271320	60	AC008758	Homo sapi
185	29	0.6	176616	60	AC007449	Homo sapi	c 258	29	0.6	324816	85	AC002368	Homo sapi
186	29	0.6	177062	61	AC009609	Homo sapi	c 259	29	0.6	327654	66	AC020816	Mus muscu
187	29	0.6	177067	64	AC016516	Homo sapi	c 260	29	0.6	340000	91	AP001710	Homo sapi
188	29	0.6	177343	65	AC018952	Homo sapi	c 261	29	0.6	340000	91	AP001728	Homo sapi
189	29	0.6	177373	83	AP002452	Homo sapi	c 262	29	0.6	340000	91	AP001754	Homo sapi
190	29	0.6	178036	82	AP001188	Homo sapi	c 263	28	0.5	284	54	G36170	STS h14a199
191	29	0.6	178109	64	AC016502	Homo sapi	c 264	28	0.5	284	54	G36170	STS h14a199
192	29	0.6	178238	64	AC016740	Homo sapi	c 265	28	0.5	511	91	BC005257	Homo sapi
193	29	0.6	178929	90	AL359252	Human DNA	c 266	28	0.5	808	12	AF022741	Oryza sat
194	29	0.6	178975	67	AC023222	Homo sapi	c 267	28	0.5	880	8	AF062745	Protocler
195	29	0.6	179671	83	AP002992	Homo sapi	c 268	28	0.5	1057	9	AX044406	Sequence
196	29	0.6	179676	62	AC012221	Homo sapi	c 269	28	0.5	1155	53	CNS067J5	AL414615 T3 end of
197	29	0.6	179929	86	AC006270	Homo sapi	c 270	28	0.5	1215	4	AB032841	Dictyoste
198	29	0.6	180163	66	AC021510	Homo sapi	c 271	28	0.5	1330	7	SSJ000928	STS scrof
199	29	0.6	180178	69	AC025287	Homo sapi	c 272	28	0.5	1432	6	HSC299386	Heteroder
200	29	0.6	180884	86	AC008018	Homo sapi	c 273	28	0.5	1436	12	AF007779	Arabidops
201	29	0.6	180942	65	AC019194	Homo sapi	c 274	28	0.5	1741	96	PMPAX6	Y09975 P.mammillat
202	29	0.6	181261	81	AL391730	Homo sapi	c 275	28	0.5	1839	89	AK022281	Homo sapi
203	29	0.6	181503	62	AC011999	Homo sapi	c 276	28	0.5	1854	91	BC005368	Homo sapi
204	29	0.6	181716	73	AC068285	Homo sapi	c 277	28	0.5	1875	97	HSU79289	Human clone
205	29	0.6	181827	67	AC022942	Homo sapi	c 278	28	0.5	3457	5	AF310886	Dictyoste
206	29	0.6	184718	87	AC018646	Homo sapi	c 279	28	0.5	3744	7	SSC251528	Dictyoste
207	29	0.6	186491	61	AC010981	Homo sapi	c 280	28	0.5	3834	14	KLKRT2GEN	AF242552 Gallus ga
208	29	0.6	186812	67	AC023089	Homo sapi	c 281	28	0.5	4103	8	AF242552	Gallus ga
209	29	0.6	187182	91	CNS01DS2	BAC seque	c 282	28	0.5	4163	6	DDIBP74	M29237 D.discoideu
210	29	0.6	187692	62	AC011303	Homo sapi	c 283	28	0.5	4617	96	SFAJ5398	Dictyoste
211	29	0.6	188396	82	AL590308	Homo sapi	c 284	28	0.5	18019	94	AB052890	Mus muscu
212	29	0.6	190000	78	AL391911	Homo sapi	c 285	28	0.5	35467	86	AC005570	Homo sapi
213	29	0.6	190844	75	AC073990	Homo sapi	c 286	28	0.5	39316	63	AC014425	Homo sapi
214	29	0.6	192535	70	AC027220	Homo sapi	c 287	28	0.5	40600	97	HSU101D3	AC014425 Drosophil
215	29	0.6	192545	62	AC012571	Homo sapi	c 288	28	0.5	43265	77	AC087378	Human DNA s
216	29	0.6	193587	61	AC009268	Homo sapi	c 289	28	0.5	43351	85	AC004800	Homo sapi
217	29	0.6	193935	83	AP002770	Homo sapi	c 290	28	0.5	44762	85	AC002325	Homo sapi
218	29	0.6	195529	69	AC025714	Homo sapi	c 291	28	0.5	45453	12	AB022212	Arabidops
219	29	0.6	195545	81	AL392090	Homo sapi	c 292	28	0.5	50149	62	AC011955	Homo sapi
220	29	0.6	196138	79	AL353576	Homo sapi	c 293	28	0.5	51737	12	AB013391	Arabidops
221	29	0.6	196554	74	AC073128	Homo sapi	c 294	28	0.5	53000	60	AC003656	Continuation (8 of
222	29	0.6	197663	87	AC009245	Human Chr	c 295	28	0.5	57196	78	AC091101	Homo sapi
223	29	0.6	197692	71	AC009554	Homo sapi	c 296	28	0.5	58238	77	AC087847	Homo sapi
224	29	0.6	198070	61	AC044909	Homo sapi	c 297	28	0.5	61373	74	AC073453	Homo sapi
225	29	0.6	198116	65	AC018440	Homo sapi	c 298	28	0.5	62850	78	AC090634	Homo sapi
226	29	0.6	198453	62	AC011359	Homo sapi	c 299	28	0.5	66260	77	AC087745	Homo sapi
227	29	0.6	198935	73	AC068573	Homo sapi	c 300	28	0.5	66594	77	AC090167	Homo sapi

9	311	5.9	54398	73	AC068224	Homo sapi	AC068224	Homo sapi	30	0.6	245624	80	AL359199	Homo sapi
10	297	5.6	54398	73	AC068224	Homo sapi	AC068224	Homo sapi	30	0.6	340000	91	AP001660	Homo sapi
11	223	4.2	72157	64	AC016449	Homo sapi	AC016449	Homo sapi	30	0.6	340000	91	AP001660	Homo sapi
12	163	3.1	113704	92	H5267P19	Z75889 Human DNA s	Z75889 Human DNA s		30	0.6	445	13	AF329891	Pisolithu
13	163	3.1	163642	79	AL353724	Homo sapi	AL353724	Homo sapi	29	0.6	1511	13	AF329891	Pisolithu
14	156	3.0	163642	79	AL353724	Homo sapi	AL353724	Homo sapi	29	0.6	2220	6	DDCAMP1	X52888 D.discoldeu
15	120	2.3	69740	77	AC087748	Homo sapi	AC087748	Homo sapi	29	0.6	2246	94	AF017152	Mus muscu
16	100	1.9	69740	77	AC087748	Homo sapi	AC087748	Homo sapi	29	0.6	4548	88	AF059683	Homo sapi
17	99	1.9	150	54	G43328	G43328 WIAF-1982-S	G43328 WIAF-1982-S		29	0.6	6372	6	DDICNPA	M23449 Dictyosellu
18	99	1.9	150	54	G43329	G43329 WIAF-1983-S	G43329 WIAF-1983-S		29	0.6	15699	85	AC003076	Human PAC
19	56	1.1	340969	81	AL512630	Mus muscu	AL512630 Mus muscu		29	0.6	15847	91	AP001061	Homo sapi
20	47	0.9	91835	92	HS266223	Z84467 Human DNA s	Z84467 Human DNA s		29	0.6	43782	93	HS0J258L1	Human DNA
21	47	0.9	153665	78	AL138818	AL138818 Homo sapi	AL138818 Homo sapi		29	0.6	44516	89	AF246630S2	Human DNA
22	46	0.9	153665	78	AL138818	AL138818 Homo sapi	AL138818 Homo sapi		29	0.6	68631	90	AL139348	Human DNA
23	41	0.8	195380	80	AL358892	AL358892 Mus muscu	AL358892 Mus muscu		29	0.6	71172	68	AC024313	Homo sapi
24	33	0.6	99030	86	AC007038	AC007038 Homo sapi	AC007038 Homo sapi		29	0.6	74236	75	AC079221	Mus muscu
25	33	0.6	115168	77	AC084283	AC084283 Homo sapi	AC084283 Homo sapi		29	0.6	77902	85	AC006958	Homo sapi
26	33	0.6	119395	80	AL139189	AL139189 Human DNA	AL139189 Human DNA		29	0.6	83549	85	AC004752	Homo sapi
27	33	0.6	177551	90	AL359436	AL359436 Homo sapi	AL359436 Homo sapi		29	0.6	85089	91	AP0001432	Homo sapi
28	32	0.6	163447	90	AL359436	AL359436 Homo sapi	AL359436 Homo sapi		29	0.6	89263	86	AC007124	Human gen
29	32	0.6	189056	68	AC024519	AC024519 Homo sapi	AC024519 Homo sapi		29	0.6	89872	93	HSJ822J19	Human DNA
30	32	0.6	239086	69	AC024946	AC024946 Homo sapi	AC024946 Homo sapi		29	0.6	93265	94	AP002736	Mus muscu
31	32	0.6	283512	80	AL365202	AL365202 Homo sapi	AL365202 Homo sapi		29	0.6	95096	86	AC008384	Homo sapi
32	31	0.6	76986	73	AC067977	AC067977 Homo sapi	AC067977 Homo sapi		29	0.6	98257	61	AC009152	Homo sapi
33	31	0.6	94409	67	AC022305	AC022305 Homo sapi	AC022305 Homo sapi		29	0.6	99469	90	AL356961	Human DNA
34	31	0.6	151088	87	AC020917	AC020917 Homo sapi	AC020917 Homo sapi		29	0.6	100000	91	AP000010	Homo sapi
35	31	0.6	156534	61	AC009700	AC009700 Homo sapi	AC009700 Homo sapi		29	0.6	100000	91	AP000131	Homo sapi
36	31	0.6	156840	86	AC006036	AC006036 Homo sapi	AC006036 Homo sapi		29	0.6	100000	91	AP000151	Homo sapi
37	31	0.6	159251	66	AC021659	AC021659 Homo sapi	AC021659 Homo sapi		29	0.6	100000	91	AP000209	Homo sapi
38	31	0.6	171444	77	AC087790	AC087790 Homo sapi	AC087790 Homo sapi		29	0.6	107484	92	HS4241L6	Human DNA
39	31	0.6	171570	70	AC026454	AC026454 Homo sapi	AC026454 Homo sapi		29	0.6	108040	73	AC068150	Homo sapi
40	31	0.6	172105	66	AC021197	AC021197 Homo sapi	AC021197 Homo sapi		29	0.6	110000	85	AC005079.2	Continuatio
41	31	0.6	175324	74	AC090890	AC090890 Canis fam	AC090890 Canis fam		29	0.6	113916	97	HUAC003007	Human Chr
42	31	0.6	179150	68	AC017077	AC017077 Homo sapi	AC017077 Homo sapi		29	0.6	114784	92	HS1189B24	Human DNA
43	31	0.6	179300	66	AC020911	AC020911 Homo sapi	AC020911 Homo sapi		29	0.6	115118	91	AP000247	Homo sapi
44	31	0.6	213999	88	AC051619	AC051619 Homo sapi	AC051619 Homo sapi		29	0.6	120359	85	AC005389	Homo sapi
45	31	0.6	245324	77	AC084794	AC084794 Homo sapi	AC084794 Homo sapi		29	0.6	122667	80	AL355998	Homo sapi
46	30	0.6	100925	83	AP002021	AP002021 Homo sapi	AP002021 Homo sapi		29	0.6	126464	92	HS225E12	Human DNA
47	30	0.6	101539	88	AC073980	AC073980 Homo sapi	AC073980 Homo sapi		29	0.6	127258	86	AC008562	Homo sapi
48	30	0.6	110042	92	HS175E3	295113 Human DNA s	295113 Human DNA s		29	0.6	128915	85	AC002416	Human Chr
49	30	0.6	116189	92	HS98L15	AC078615 Homo sapi	AC078615 Homo sapi		29	0.6	129366	92	HS224A6	Human Chr
50	30	0.6	128822	82	AP001260	AP001260 Homo sapi	AP001260 Homo sapi		29	0.6	130659	74	AC069368	Homo sapi
51	30	0.6	136769	91	AC026453	AC026453 Homo sapi	AC026453 Homo sapi		29	0.6	130831	70	AC027568	Homo sapi
52	30	0.6	136804	91	AP000469	AP000469 Homo sapi	AP000469 Homo sapi		29	0.6	134253	64	AC015923	Homo sapi
53	30	0.6	139391	86	AC007239	AC007239 Homo sapi	AC007239 Homo sapi		29	0.6	134978	75	AC079004	Homo sapi
54	30	0.6	143839	85	AC004701	AC004701 Homo sapi	AC004701 Homo sapi		29	0.6	137951	82	AL590101	Homo sapi
55	30	0.6	147519	75	AC079099	AC079099 Homo sapi	AC079099 Homo sapi		29	0.6	138056	93	HSJ876B10	Human DNA
56	30	0.6	152706	79	AL355377	AL355377 Homo sapi	AL355377 Homo sapi		29	0.6	139229	78	AL139156	Homo sapi
57	30	0.6	155450	62	AC012324	AC012324 Homo sapi	AC012324 Homo sapi		29	0.6	139768	61	AC010931	Homo sapi
58	30	0.6	155722	72	AC053487	AC053487 Homo sapi	AC053487 Homo sapi		29	0.6	141172	91	AL512600	Human DNA
59	30	0.6	156833	86	AC006556	AC006556 Homo sapi	AC006556 Homo sapi		29	0.6	143291	89	AL137792	Human DNA
60	30	0.6	158037	62	AC012126	AC012126 Homo sapi	AC012126 Homo sapi		29	0.6	143697	80	AL358572	Homo sapi
61	30	0.6	159720	88	AC068716	AC068716 Homo sapi	AC068716 Homo sapi		29	0.6	146158	89	AL136136	Human DNA
62	30	0.6	162584	83	AP002763	AP002763 Homo sapi	AP002763 Homo sapi		29	0.6	146329	78	AC090889	Canis fam
63	30	0.6	165120	89	AL138704	AL138704 Human DNA	AL138704 Human DNA		29	0.6	146335	63	AC012626	Homo sapi
64	30	0.6	166782	79	AL353896	AL353896 Homo sapi	AL353896 Homo sapi		29	0.6	147086	85	AC002472	Homo sapi
65	30	0.6	168305	69	AC024969	AC024969 Homo sapi	AC024969 Homo sapi		29	0.6	148069	93	HSJ86016	Homo sapi
66	30	0.6	168440	79	AL162572	AL162572 Homo sapi	AL162572 Homo sapi		29	0.6	150355	93	HSJ864H10	Human DNA
67	30	0.6	168586	87	AC011497	AC011497 Homo sapi	AC011497 Homo sapi		29	0.6	150424	90	AL391811	Human DNA
68	30	0.6	169413	80	AL358118	AL358118 Homo sapi	AL358118 Homo sapi		29	0.6	150697	69	AC025423	Homo sapi
69	30	0.6	175304	60	AC008167	AC008167 Homo sapi	AC008167 Homo sapi		29	0.6	150715	64	AC016258	Homo sapi
70	30	0.6	178169	90	AL391137	AL391137 Human DNA	AL391137 Human DNA		29	0.6	152452	77	AC090105	Homo sapi
71	30	0.6	188679	67	AC023121	AC023121 Homo sapi	AC023121 Homo sapi		29	0.6	152742	70	AC027005	Homo sapi
72	30	0.6	189978	87	AC009075	AC009075 Homo sapi	AC009075 Homo sapi		29	0.6	152794	90	AL359085	Human DNA
73	30	0.6	190007	79	AL158155	AL158155 Homo sapi	AL158155 Homo sapi		29	0.6	153206	86	AC006356	Homo sapi
74	30	0.6	190793	71	AC034298	AC034298 Homo sapi	AC034298 Homo sapi		29	0.6	154835	73	AC069045	Homo sapi
75	30	0.6	190963	68	AC023597	AC023597 Homo sapi	AC023597 Homo sapi		29	0.6	155391	76	AC083801	Homo sapi
76	30	0.6	194520	90	AL356915	AL356915 Human DNA	AL356915 Human DNA		29	0.6	155414	74	AC069176	Homo sapi
77	30	0.6	198829	81	AL391668	AL391668 Homo sapi	AL391668 Homo sapi		29	0.6	156100	62	AC011934	Homo sapi
78	30	0.6	199489	70	AC027505	AC027505 Homo sapi	AC027505 Homo sapi		29	0.6	156100	62	AC011934	Homo sapi
79	30	0.6	200000	60	AC004604	AC004604 Homo sapi	AC004604 Homo sapi		29	0.6	157485	75	AC079056	Homo sapi
80	30	0.6	203748	74	AC069250	AC069250 Homo sapi	AC069250 Homo sapi		29	0.6	158427	85	AC002553	Homo sapi
81	30	0.6	213632	78	AL136145	AL136145 Homo sapi	AL136145 Homo sapi		29	0.6	159126	75	AC078795	Homo sapi

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:46:47 ; Search time 6800.24 Seconds
(without alignments)
11989.357 Million cell updates/sec

Title: US-09-512-581-1
Perfect score: 5271
Sequence: 1 ccggagagcccgagtgag.....gtaaaaaaaaaaaaaa 5271

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 1344157 seqs, 7733874588 residues

Word size: 0
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl.*			
1:	gb_ba1.*	44:	em_ov.*
2:	gb_ba2.*	45:	em_pat.*
3:	gb_ba3.*	46:	em_ph.*
4:	gb_in1.*	47:	em_pl.*
5:	gb_in2.*	48:	em_ro.*
6:	gb_in3.*	49:	em_sts.*
7:	gb_om.*	50:	em_sy.*
8:	gb_ov.*	51:	em_un.*
9:	gb_pat1.*	52:	em_vi.*
10:	gb_pat2.*	53:	gb_sts1.*
11:	gb_ph.*	54:	gb_sts2.*
12:	gb_pl1.*	55:	gb_sts3.*
13:	gb_pl2.*	56:	gb_sy.*
14:	gb_pl3.*	57:	gb_un.*
15:	gb_pl4.*	58:	gb_vil.*
16:	gb_ba1.*	59:	gb_vil2.*
17:	em_ba2.*	60:	gb_htg1.*
18:	em_fun.*	61:	gb_htg2.*
19:	em_htgo_hum.*	62:	gb_htg3.*
20:	em_htgo_inv.*	63:	gb_htg4.*
21:	em_htgo_rod.*	64:	gb_htg5.*
22:	em_htg_hum1.*	65:	gb_htg6.*
23:	em_htg_hum2.*	66:	gb_htg7.*
24:	em_htg_hum3.*	67:	gb_htg8.*
25:	em_htg_hum4.*	68:	gb_htg9.*
26:	em_htg_hum5.*	69:	gb_htg10.*
27:	em_htg_hum6.*	70:	gb_htg11.*
28:	em_htg_hum7.*	71:	gb_htg12.*
29:	em_htg_hum8.*	72:	gb_htg13.*
30:	em_htg_inv1.*	73:	gb_htg14.*
31:	em_htg_inv2.*	74:	gb_htg15.*
32:	em_htg_other.*	75:	gb_htg16.*
33:	em_hum1.*	76:	gb_htg17.*
34:	em_hum2.*	77:	gb_htg18.*
35:	em_hum3.*	78:	gb_htg19.*
36:	em_hum4.*	79:	gb_htg20.*
37:	em_hum5.*	80:	gb_htg21.*
38:	em_hum6.*	81:	gb_htg22.*
39:	em_hum7.*	82:	gb_htg23.*
40:	em_in.*	83:	gb_htg24.*
41:	em_in.*	84:	gb_htg25.*
42:	em_om.*	85:	gb_pr1.*
43:	em_or.*	86:	gb_pr2.*
		87:	gb_pr3.*
		88:	gb_pr4.*
		89:	gb_pr5.*
		90:	gb_pr6.*
		91:	gb_pr7.*
		92:	gb_pr8.*
		93:	gb_pr9.*
		94:	gb_rol.*
		95:	gb_rol2.*
		96:	gb_in4.*
		97:	gb_pr10.*
		98:	em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	80.3	5253	97	HS095825
2	4033	76.5	7444	92	HS26H231
3	3826	72.6	5110	85	AB023196
4	1126	21.4	1852	97	HS050533
5	850	16.1	137246	92	HS49J10
6	760	14.4	168487	89	AL138820
7	441	8.4	2079	89	AK026889
8	441	8.4	72157	64	AC016449

THIS PAGE BLANK (USPTO)

CC -!- PTM: COVALENTLY LINKED TO A SMALL RNA OF ABOUT 20 NUCLEOTIDES.
DR PIR: A01520; TNBOA1.
DR PIR: S06426; S06426.
KW Immunopotentialization; Thymus; Acetylation; Nuclear protein;
KW Phosphorylation.
FT PEPTIDE 1 28 THYMOSIN ALPHA-1.
FT MOD_RES 1 1 ACETYLATION.
FT DOMAIN 40 99 ASP/GLU-RICH (ACIDIC).
FT MOD_RES 7 7 PHOSPHORYLATION (BY CK2).
FT MOD_RES 12 12 PHOSPHORYLATION (BY CK2).
FT MOD_RES 13 13 PHOSPHORYLATION (BY CK2).
SQ SEQUENCE 109 AA; 11941 MW; ECB151F0A9F99D24 CRC64;

Query Match 0.5%; Score 7; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1331
DB 60 EEEEEEE 66

RESULT 50
THYA_HUMAN
ID THYA_HUMAN STANDARD; PRT; 110 AA.
AC P06454;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTHYMOSIN ALPHA.
GN PTMA OR TWSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=87092260; PubMed=3467312;
RA Eschenfeldt W.H., Berger S.L.;
RT "The human prothymosin alpha gene is polymorphic and induced upon
RT growth stimulation: evidence using a cloned cDNA";
RT Proc. Natl. Acad. Sci. U.S.A. 83:9403-9407(1986).
RN [2]
RS SEQUENCE FROM N.A.
RX MEDLINE=87067426; PubMed=3466166;
RA Goodall G.J., Dominguez F., Horecker B.L.;
RT "Molecular cloning of cDNA for human prothymosin alpha";
RT Proc. Natl. Acad. Sci. U.S.A. 83:8926-8928(1986).
RN [3]
RS SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89214202; PubMed=2708378;
RA Eschenfeldt W.H., Manrow R.E., Berger S.L.;
RT "Isolation and partial sequencing of the human prothymosin alpha gene
RT family. Evidence against export of the gene products.";
RT J. Biol. Chem. 264:7546-7555(1989).
RN [4]
RS SEQUENCE OF 1-14 FROM N.A.
RX MEDLINE=93186133; PubMed=7916742;
RA Szabo P., Panneerselvam C., Clinton M., Frangou-Lazaridis M.,
RA Weksler D., Whittington E., Macera M.J., Grzeschik K.H.,
RA Selvakumar A., Horecker B.L.;
RT "Prothymosin alpha gene in humans: organization of its promoter
RT region and localization to chromosome 2";
RT Hum. Genet. 90:629-634(1993).
CC -!- FUNCTION: PROTHYMOSIN ALPHA MAY MEDIATE IMMUNE FUNCTION BY
CC CONFERRING RESISTANCE TO CERTAIN OPPORTUNISTIC INFECTIONS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- PTM: COVALENTLY LINKED TO A SMALL RNA OF ABOUT 20 NUCLEOTIDES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67480; AAA63240.1; -
DR EMBL; J04797; AAA63240.1; JOINED.
DR EMBL; M67480; AAA63239.1; -
DR EMBL; J04797; AAA63239.1; JOINED.
DR EMBL; M14483; AAA61183.1; -
DR EMBL; M14630; AAA61182.1; -
DR EMBL; S56449; AAD13882.1; -
DR PIR; A25920; TNHUA.
DR PIR; A26654; A26654.
DR PIR; B33356; B33356.
DR MIM; 188390; -
KW Immunopotentialization; Thymus; Acetylation; Nuclear protein.
FT INIT_MET 0 0 THYMOSIN ALPHA-1.
FT PEPTIDE 1 28 ACETYLATION (PROBABLE).
FT MOD_RES 1 1 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 41 100 MISSING (IN REF. 2).
FT CONFLICT 39 39
SQ SEQUENCE 110 AA; 12072 MW; 0A6609469986D272 CRC64;

Query Match 0.5%; Score 7; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1331
DB 61 EEEEEEE 67

Search completed: September 25, 2001, 10:08:03
Job time: 154 sec


```

OX NCBI_TaxID=10116;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=87277429; PubMed=2886401;
RA Falzon M., Sanderson N., Chung S.X.;
RT "Cloning and expression of rat homeo-box-containing sequences.";
RL Gene 54:23-32(1987).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16807; -; NOT_ANNOTATED_CDS.
DR PIR; A27471; A27471.
DR HSSP; P02833; 1HOM.
DR TRANSFAC; T01707; -.
DR InterPro; IPR001356; -.
DR InterPro; IPR001827; -.
DR Pfam; PF00046; homeobox; 1.
DR PROSITE; PS00032; ANTENNAPEDIA; PARTIAL.
DR PROSITE; PS00027; HOMEBOX.1; 1.
DR PROSITE; PS00071; HOMEBOX.2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT NON_TER 1 1
FT DNA_BIND 5 64 HOMEBOX.
FT DOMAIN 91 105 GLU-RICH (ACIDIC).
FT SEQUENCE 105 AA; 12552 MW; 106C1DF938F2864B CRC64;
SQ
-----
Query Match 0.5%; Score 7; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 91 EEEEEEE 97
-----
RESULT 48
ID RL12_ARCFU STANDARD; PRT; 106 AA.
AC O28780;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L12P.
GN RPL12P OR APL1492.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervlavage A.R., Graham D.E., Kyrpides N.C.,
RA Flikschmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

```

```

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000999; AAB89748.1; -.
DR TIGR; AFL1492; -.
DR InterPro; IPR001813; -.
DR Pfam; PF00428; 60s_ribosomal; 1.
DR Ribosomal protein.
KW Ribosomal protein.
SQ SEQUENCE 106 AA; 11037 MW; 48956678256633B1 CRC64;
-----
Query Match 0.5%; Score 7; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 84 EEEEEEE 90
-----
RESULT 49
ID THYA_BOVIN STANDARD; PRT; 109 AA.
AC P01252;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTHYMOSIN ALPHA.
GN PTMA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RE SEQUENCE.
RX MEDLINE=88339364; PubMed=2901823;
RA Panneerselvam C., Wellner D., Horecker B.L.;
RT "The amino acid sequence of bovine thymus prothymosin alpha.";
RL Arch. Biochem. Biophys. 265:454-457(1988).
RN [2]
RE SEQUENCE OF 1-28.
RX MEDLINE=79109664; PubMed=762108;
RA Low T.L.K., Goldstein A.L.;
RT "The chemistry and biology of thymosin. II. Amino acid sequence
RT analysis of thymosin alpha and polypeptide beta.";
RL J. Biol. Chem. 254:987-995(1979).
RN [3]
RE PHOSPHORYLATION BY CK2.
RX MEDLINE=93050183; PubMed=1426245;
RA Barcia M.G., Castro J.M., Jullien C.D., Gonzalez C.G., Freire M.;
RT "Prothymosin alpha is phosphorylated by casein kinase-2.";
RL FEBS Lett. 312:152-156(1992).
CC -!- FUNCTION: PROTHYMOSIN ALPHA MAY MEDIATE IMMUNE FUNCTION BY
CC CONFERRING RESISTANCE TO CERTAIN OPPORTUNISTIC INFECTIONS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC

```


RA MEDLINE-89338759; PubMed-2759245;
RA Trompeter H.-I., Brand I.A., Soeling H.-D.;
RT "The primary sequence of the PFK-1 inactivating zinc-binding protein
RT as deduced from cDNA sequencing. Identity of the zinc-binding protein
RT with rat parathymosin.";
RL FEBS Lett. 253:63-66(1989).
RN [2]
RF SEQUENCE FROM N.A.
RC STRAIN-FISCHER; TISSUE=Liver;
RX MEDLINE-92183876; PubMed-1544455;
RA Trompeter H.-I., Soeling H.-D.;
RT "Cloning and characterization of a gene encoding the 11.5 kDa zinc-
RT binding protein (parathymosin-alpha).";
RL FEBS Lett. 298:245-248(1992).
RN [3]
RF SEQUENCE OF 22-101 FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE-88239673; PubMed-3377505;
RA Prangou-Lazaridis M., Clinton M., Goodall G.J., Horecker B.L.;
RT "Prothymosin alpha and parathymosin: amino acid sequences deduced
RT from the cloned rat spleen cDNAs.";
RL Arch. Biochem. Biophys. 263:305-310(1988).
RN [4]
RF PRELIMINARY SEQUENCE
RC MEDLINE-86149281; PubMed-3456585;
RA Komiyama T., Pan L.-X., Haritos A.A., Wideman J.W., Pan Y.-C.E.,
RT Chang M., Rogers I., Horecker B.L.;
RT "The primary structure of rat parathymosin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1242-1245(1986).
RN [5]
RF PRELIMINARY SEQUENCE OF 1-30.
RC MEDLINE-85140226; PubMed-3856246;
RA Haritos A.A., Salvin S.B., Blacher R., Stein S., Horecker B.L.;
RT "Parathymosin alpha: a peptide from rat tissues with structural
RT homology to prothymosin alpha.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1050-1053(1985).
RN [6]
RF REVISION TO 1-4.
RC MEDLINE-88339930; PubMed-3421960;
RA Panneerselvam C., Clinton M., Wellner D., Horecker B.L.;
RT "Bovine parathymosin: amino acid sequence and comparison with rat
RT parathymosin.";
RL Biochem. Biophys. Res. Commun. 155:539-545(1988).
CC -!- FUNCTION: PARATHYMOSIN MAY MEDIATE IMMUNE FUNCTION BY BLOCKING
CC CERTAIN OPPORTUNISTIC INFECTIONS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20616; AAA42249.1; -
CC EMBL; M33025; AAA41810.1; -
CC EMBL; X64053; CAA45411.1; ALT_SEQ.
CC EMBL; X16481; CAA34501.1; -
CC PIR; B31512; B31512
CC PIR; A05095; A05095.
CC PIR; A22075; A22075.
CC PIR; B32265; B32265.
CC PIR; S05212; S05212.
KW Immunopotentialization; Thymus; Acetylation; Zinc.
FT INIT_MFT 0
FT MOD_RES 1
FT MOD_RES 35
FT DOMAIN 1 89
FT DOMAIN ASP/GLU-RICH (ACIDIC).
FT SEQUENCE 101 AA; 11428 MW; D32BE03032B85D4 CRC64;
Query Match 0.5%; Score 7; DB 1; Length 101;
Best Local Similarity 100.08; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 66 EEEEEEE 72
RESULT 46
HSP3_MOUSE
ID HSP3_MOUSE STANDARD; PRT; 102 AA.
AC Q62100;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P3.
GN PRM3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE-96341725; PubMed-8720108;
RA Schluter G., Celik A.B., Obata R., Schlicker M., Hofferbert S.,
RT Schlung A., Adham I.M., Engel W.;
RT "Sequence analysis of the conserved protamine gene cluster shows that
RT it contains a fourth expressed gene.";
RL Mol. Reprod. Dev. 43:1-6(1996).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 247352; CAA87412.1; -
CC MGD; MGI:106601; Prm3.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT DOMAIN 46 68
FT DOMAIN ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 75 78
FT DOMAIN POLY-LEU.
FT SEQUENCE 102 AA; 11409 MW; E7AA59D3313AF846C CRC64;
Query Match 0.5%; Score 7; DB 1; Length 102;
Best Local Similarity 100.08; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 54 EEEEEEE 60
RESULT 47
HXA7_RAT
ID HXA7_RAT STANDARD; PRT; 105 AA.
AC P09634;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HOMEOBOX PROTEIN HOXA-7 (HOX-1.1) (R5) (FRAGMENT).
GN HOXA7 OR HOXA-7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

CC PROXIMAL SEQUENCE ELEMENT (PSE), A NON-TATA-BOX BASAL PROMOTER
CC ELEMENT COMMON TO THESE 2 TYPES OF GENES.
CC -1- SUBUNIT: COMPOSED OF 5 SUBUNITS: SNAP190; SNAP50; SNAP45; SNAP43
CC AND SNAP19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF093593; AAC61873.1; .
CC Transcription regulation. POLY-GLU.
CC DOMAIN 86 95
CC SEQUENCE 98 AA; 11328 MW; 4D797E35AF2D1485 CRC64;
CC -----
CC Query Match 0.5%; Score 7; DB 1; Length 98;
CC Best Local Similarity 100.0%; Pred. No. 39;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1325 EEEEEEE 1331
CC |
CC DB 86 EEEEEEE 92
CC -----
CC RESULT 43
CC RLI2_METH
CC ID RLI2_METH STANDARD; PRT; 101 AA.
CC AC P05394;
CC DT 01-NOV-1988 (Rel. 09, Created)
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE 50S RIBOSOMAL PROTEIN L12P ('A' TYPE).
CC GN RPL12P OR MTH1682.
CC OS Methanobacterium thermoautotrophicum.
CC OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
CC OC Methanothermobacter.
CC OX NCBI_TaxID=145262;
CC RN [1]
CC RE SEQUENCE FROM N.A.
CC RC STRAIN=DELTA H;
CC RX MEDLINE=98037514; PubMed=9371463;
CC RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
CC RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
CC RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
CC RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
CC RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
CC RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
CC RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
CC RT "Complete genome sequence of Methanobacterium thermoautotrophicum
CC RT deltaH: functional analysis and comparative genomics.";
CC RU J. Bacteriol. 179:7135-7155(1997).
CC RN [2]
CC RE SEQUENCE OF 1-48.
CC RX MEDLINE=81110590; PubMed=7006702;
CC RA Matheson A.T., Yaguchi M., Balch W.E., Wolfe R.S.;
CC RT "Sequence homologies in the N-terminal region of the ribosomal 'A'
CC RT proteins from Methanobacterium thermoautotrophicum and Halobacterium
CC RT cutirubrum.";
CC RU Biochim. Biophys. Acta 626:162-169(1980).
CC CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION BY SIMILARITY.
CC CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000926; AAB86154.1; .
CC InterPro; IPR001813; .
CC DR Pfam; PF00428; 60s_ribosomal; 1.
CC KW Ribosomal protein.
CC SQ SEQUENCE 101 AA; 10480 MW; 6C1CEA3EF121PB85 CRC64;
CC -----
CC Query Match 0.5%; Score 7; DB 1; Length 101;
CC Best Local Similarity 100.0%; Pred. No. 40;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1325 EEEEEEE 1331
CC |
CC DB 80 EEEEEEE 86
CC -----
CC RESULT 44
CC THYP_BOVIN
CC ID THYP_BOVIN STANDARD; PRT; 101 AA.
CC AC P08814;
CC DT 01-NOV-1988 (Rel. 09, Created)
CC DT 01-NOV-1988 (Rel. 09, Last sequence update)
CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
CC DE PARATHYMOSIN.
CC GN PTMS.
CC OS Bos taurus (Bovine).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC OC Bovidae; Bovinae; Bos.
CC OX NCBI_TaxID=9913;
CC RN [1]
CC RE SEQUENCE.
CC RX MEDLINE=88339930; PubMed=3421960;
CC RA Panneerselvam C., Clinton M., Wellner D., Horecker B.L.;
CC RT "Bovine parathymosin: amino acid sequence and comparison with rat
CC RT parathymosin.";
CC RL Biochem. Biophys. Res. Commun. 155:539-545(1988).
CC CC -1- FUNCTION: PARATHYMOSIN MAY MEDIATE IMMUNE FUNCTION BY BLOCKING
CC THE EFFECT OF PROTHYMOSIN ALPHA WHICH CONFERS RESISTANCE TO
CC CERTAIN OPPORTUNISTIC INFECTIONS.
CC KW Immunopotentialization; Thymus; Acetylation.
CC FT MOD_RES 1 1 ACETYLATION.
CC FT DOMAIN 35 89 ASP/GLU-RICH (ACIDIC).
CC SQ SEQUENCE 101 AA; 11326 MW; 187F7D05B603000C CRC64;
CC -----
CC Query Match 0.5%; Score 7; DB 1; Length 101;
CC Best Local Similarity 100.0%; Pred. No. 40;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1325 EEEEEEE 1331
CC |
CC DB 66 EEEEEEE 72
CC -----
CC RESULT 45
CC THYP_RAT
CC ID THYP_RAT STANDARD; PRT; 101 AA.
CC AC P04550;
CC DT 13-AUG-1987 (Rel. 05, Created)
CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE PARATHYMOSIN (ZINC-BINDING 11.5 KDA PROTEIN).
CC GN PTMS OR ZNBP.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC OX NCBI_TaxID=10116;
CC RN [1]
CC RE SEQUENCE FROM N.A.


```
CC 2 FERROCYTOCHROME C.
CC -1- SUBUNIT: BCL COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS
CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
CC 6 LOW-MOLECULAR WEIGHT PROTEINS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE UOCHR/OCR6 FAMILY.
DR PIR: A00119; CCB011.
DR PDB: 1QCR; 14-OCT-98.
DR PDB: 1BGY; 23-FEB-99.
DR PDB: 1BE3; 16-FEB-99.
KW Mitochondrion: Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; 3D-structure. POLY-GLU.
FT DOMAIN 5
FT SEQUENCE 78 AA; 9175 MW; 4A2D25A555455560 CRC64;
SQ
-----
Query Match 0.5%; Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 5 EEEEEEE 11
-----
RESULT 40
EHL1_ADEM1
ID EHL1_ADEM1 STANDARD; PRT; 96 AA.
AC P12533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE EARLY E1A 11 KDA PROTEIN.
OS Mouse adenovirus type 1 (MAV-1).
OC Viruses: dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10530;
RN [1]
RF SEQUENCE FROM N.A.
RX MEDLINE=89268474; PubMed=2543128;
RA Ball A.O., Beard C.W., Redick S.D., Spindler K.R.;
RT "Genome organization of mouse adenovirus type 1 early region 1: a
RT novel transcription map.";
RL Virology 170:523-536(1989).
RN [2]
RF SEQUENCE OF 1-66 FROM N.A.
RX MEDLINE=89012162; PubMed=3172335;
RA Ball A.O., Williams M.E., Spindler K.R.;
RT "Identification of mouse adenovirus type 1 early region 1: DNA
RT sequence and a conserved transactivating function.";
RL J. Virol. 62:3947-3957(1988).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M22245; AAA42430.1; -.
Early protein.
KW SEQUENCE 96 AA; 11133 MW; B3CD4C1332F5547E CRC64;
SQ
-----
Query Match 0.5%; Score 7; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 76 EEEEEEE 82
-----
```

```
RESULT 41
VE7_COPY
ID VE7_COPY STANDARD; PRT; 97 AA.
AC Q89759;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE E7 PROTEIN.
GN E7.
OS Canine oral papillomavirus (COPV).
OC Viruses: dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
OX NCBI_TaxID=35258;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=Y62;
RA Isegawa N., Ohta M., Shirasawa H., Tokita H., Simizu B., Yamaura A.;
RT "Nucleotide sequence of a canine oral papillomavirus containing a long
RT noncoding region.";
RL Int. J. Oncol. 7:155-159(1995).
CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING
CC ACTIVITIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; D55633; BAA09499.1; -.
DR EMBL; L22695; AAA61745.1; -.
DR InterPro; IPR000148; -.
DR Pfam; PF00527; E7: 1.
KW Early protein; Transcription regulation; Transforming protein;
KW DNA-binding; Trans-acting factor.
FT DOMAIN 52 55 C-XX-C MOTIF.
FT DOMAIN 84 87 C-XX-C MOTIF.
SQ SEQUENCE 97 AA; 10798 MW; 7E3102AFA49B2A62 CRC64;
-----
Query Match 0.5%; Score 7; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEE 1331
DB 33 EEEEEEE 39
-----
RESULT 42
SNCS5_HUMAN
ID SNCS5_HUMAN STANDARD; PRT; 98 AA.
AC Q75971;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SNRNA ACTIVATING PROTEIN COMPLEX 19 KDA SUBUNIT (SNAPC 19 KDA
DE SUBUNIT).
GN SNAPC5 OR SNAP19
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RF SEQUENCE FROM N.A.
RC MEDLINE=98402477; PubMed=9732265;
RA Henry R.W., Mittal V., Ma B., Kobayashi R., Hernandez N.;
RT "SNAPc19 mediates the assembly of a functional core promoter complex
RT (SNAPc) shared by RNA polymerases II and III.";
RL Genes Dev. 12:2664-2672(1998).
CC -1- FUNCTION: COMPLEX REQUIRED FOR THE TRANSCRIPTION OF BOTH RNA
CC POLYMERASE II AND III SMALL-NUCLEAR RNA GENES. BINDS TO THE
```


[1] SEQUENCE FROM N.A.
MEDLINE=97123494; PubMed=8968741;
Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,
Gibbons R.J.;
"ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.";
Hum. Mol. Genet. 5:1899-1907(1996).
[2] SEQUENCE FROM N.A.
MEDLINE=97386582; PubMed=9244431;
Villard L., Lossi A.M., Cardoso C., Proud V., Chiaroni P.,
Colleaux L., Schwartz C., Fontes M.;
"Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase.";
Genomics 43:149-155(1997).
[3] SEQUENCE OF 743-2375 FROM N.A.
MEDLINE=95179111; PubMed=7874112;
Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,
Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
Bianchi M.E., Consalez G.G.;
"Cloning and characterization of a new human Xq13 gene, encoding a putative helicase.";
Hum. Mol. Genet. 3:1957-1964(1994).
[4] PRELIMINARY PARTIAL SEQUENCE FROM N.A.
MEDLINE=94214473; PubMed=8162050;
Gecz J., Pollard H., Consalez G., Villard L., Stayton C.,
Millasseau P., Khrestchatsky M., Fontes M.;
"Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
Hum. Mol. Genet. 3:39-44(1994).
[5] SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.
MEDLINE=95211835; PubMed=7697114;
Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
"Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
Cell 80:837-845(1995).
[6] VARIANTS ATR-X.
MEDLINE=97467722; PubMed=9326931;
Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
Higgs D.R.;
"Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain.";
Nat. Genet. 17:146-148(1997).
[7] VARIANT JM GLN-2014.
MEDLINE=96224392; PubMed=8630485;
Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-veber P.,
Munnich A., Lyonnet S.;
"XNP mutation in a large family with Juberger-Marsidi syndrome.";
Nat. Genet. 12:359-360(1996).
-!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
-!- SUBCELLULAR LOCATION: NUCLEAR.
-!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL ABNORMALITIES, AND ALPHA-THALASSEMIA.
-!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS, MICROGENITALISM AND EARLY DEATH.
-!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
-!- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U72938; AAB49971.1; -
DR EMBL; U72935; AAB40699.1; -
DR EMBL; U72904; AAB40699.1; JOINED.
DR EMBL; U72907; AAB40699.1; JOINED.
DR EMBL; U72908; AAB40699.1; JOINED.
DR EMBL; U72909; AAB40699.1; JOINED.
DR EMBL; U72910; AAB40699.1; JOINED.
DR EMBL; U72911; AAB40699.1; JOINED.
DR EMBL; U72912; AAB40699.1; JOINED.
DR EMBL; U72913; AAB40699.1; JOINED.
DR EMBL; U72914; AAB40699.1; JOINED.
DR EMBL; U72915; AAB40699.1; JOINED.
DR EMBL; U72916; AAB40699.1; JOINED.
DR EMBL; U72918; AAB40699.1; JOINED.
DR EMBL; U72919; AAB40699.1; JOINED.
DR EMBL; U72920; AAB40699.1; JOINED.
DR EMBL; U72921; AAB40699.1; JOINED.
DR EMBL; U72922; AAB40699.1; JOINED.
DR EMBL; U72923; AAB40699.1; JOINED.
DR EMBL; U72924; AAB40699.1; JOINED.
DR EMBL; U72925; AAB40699.1; JOINED.
DR EMBL; U72926; AAB40699.1; JOINED.
DR EMBL; U72927; AAB40699.1; JOINED.
DR EMBL; U72928; AAB40699.1; JOINED.
DR EMBL; U72929; AAB40699.1; JOINED.
DR EMBL; U72930; AAB40699.1; JOINED.
DR EMBL; U72931; AAB40699.1; JOINED.
DR EMBL; U72932; AAB40699.1; JOINED.
DR EMBL; U72933; AAB40699.1; JOINED.
DR EMBL; U72934; AAB40699.1; JOINED.
DR EMBL; U72935; AAB49969.1; -
DR EMBL; U72938; AAB40700.1; JOINED.
DR EMBL; U72908; AAB40700.1; JOINED.
DR EMBL; U72909; AAB40700.1; JOINED.
DR EMBL; U72910; AAB40700.1; JOINED.
DR EMBL; U72911; AAB40700.1; JOINED.
DR EMBL; U72912; AAB40700.1; JOINED.
DR EMBL; U72913; AAB40700.1; JOINED.
DR EMBL; U72914; AAB40700.1; JOINED.
DR EMBL; U72915; AAB40700.1; JOINED.
DR EMBL; U72916; AAB40700.1; JOINED.
DR EMBL; U72917; AAB40700.1; JOINED.
DR EMBL; U72918; AAB40700.1; JOINED.
DR EMBL; U72920; AAB40700.1; JOINED.
DR EMBL; U72921; AAB40700.1; JOINED.
DR EMBL; U72922; AAB40700.1; JOINED.
DR EMBL; U72923; AAB40700.1; JOINED.
DR EMBL; U72924; AAB40700.1; JOINED.
DR EMBL; U72925; AAB40700.1; JOINED.
DR EMBL; U72926; AAB40700.1; JOINED.
DR EMBL; U72927; AAB40700.1; JOINED.
DR EMBL; U72928; AAB40700.1; JOINED.
DR EMBL; U72929; AAB40700.1; JOINED.
DR EMBL; U72930; AAB40700.1; JOINED.
DR EMBL; U72931; AAB40700.1; JOINED.
DR EMBL; U72932; AAB40700.1; JOINED.
DR EMBL; U72933; AAB40700.1; JOINED.
DR EMBL; U72934; AAB40700.1; JOINED.
DR EMBL; X83753; CAA58711.1; -
DR EMBL; U97103; AAC51657.1; -
DR EMBL; AF000157; AAC51657.1; JOINED.
DR EMBL; AF000158; AAC51657.1; JOINED.
DR EMBL; AF000159; AAC51657.1; JOINED.
DR EMBL; AF000160; AAC51657.1; JOINED.
DR EMBL; U97080; AAC51657.1; JOINED.


```

EMBL; M32698; AAA41659.1; -.
DR PIR; S06006; S06006.
DR PIR; A02989; A02989.
DR HSP; P08799; LMND.
DR InterPro; IPR000048; -.
DR InterPro; IPR001609; -.
DR InterPro; IPR002928; -.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 839
FT DOMAIN 840 1935
FT DOMAIN 780 802
FT DOMAIN 840 1935
FT NP_BIND 178 185
FT DOMAIN 655 677
FT DOMAIN 757 771
FT MOD_RES 129 129
FT MOD_RES 695 695
FT MOD_RES 705 705
FT CONFLICT 1529 1531
FT CONFLICT 1731 1731
FT CONFLICT 1784 1784
FT CONFLICT 1851 1851
FT CONFLICT 1858 1858
FT CONFLICT 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;
SQ
SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1935;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKQLE 564
DB 1527 EKIRKQLE 1534

RESULT 36
CYAL_DROME STANDARD; PRT; 2248 AA.
AC P32870;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CA2+/CALMODULIN-RESPONSIVE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (RUTABAGA PROTEIN).
DE RUT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=Head;
RX MEDLINE=92154664; PubMed=1739965;
RA Levin L.R., Han P.-L., Hwang P.M., Feinstein P.G., Davis R.L.,
RA Reed R.R.;
RT "The Drosophila learning and memory gene rutabaga encodes a
RT Ca2+/calmodulin-responsive adenylyl cyclase."
RT Cell 68:479-489(1992).
CC -!- FUNCTION: THIS A MEMBRANE-BOUND, CALMODULIN-SENSITIVE ADENYLYL
CC CYCLASE. INACTIVATION OF THIS CYCLASE LEADS TO A LEARNING AND
CC MEMORY DEFECT.
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -!- ENZYME REGULATION: ACTIVATED BY CA(2+)/CALMODULIN AND G PROTEIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MUSHROOM BODIES OF THE FLY BRAIN.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.

-!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M81887; AAA28844.1; -.
DR PIR; D42088; D42088.
DR HSP; P19754; IAWK.
DR FlyBase; FBgn0003301; rut.
DR InterPro; IPR001054; -.
DR Pfam; PF00211; guanylate_cyc; 2.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
DR PROSITE; PS01125; GUANYLATE_CYCLASES_2; 2.
KW Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
FT DOMAIN 1 41
FT TRANSMEM 42 60
FT TRANSMEM 65 84
FT TRANSMEM 101 115
FT TRANSMEM 122 142
FT TRANSMEM 152 174
FT TRANSMEM 186 206
FT TRANSMEM 207 705
FT TRANSMEM 706 726
FT TRANSMEM 730 750
FT TRANSMEM 770 791
FT TRANSMEM 792 813
FT TRANSMEM 814 834
FT TRANSMEM 842 867
FT TRANSMEM 889 2248
FT DOMAIN 318 334
FT DOMAIN 1013 1029
FT DOMAIN 515 530
FT DOMAIN 569 602
FT DOMAIN 1278 1297
FT DOMAIN 1767 1810
FT DOMAIN 2025 2040
FT DOMAIN 2200 2241
FT CARBOHYD 800 800
FT CARBOHYD 807 807
FT MUTAGEN 1026 1026
SQ SEQUENCE 2248 AA; 248899 MW; E459C718BE018868 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 2248;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 LLLSVLPQ 276
DB 238 LLLSVLPQ 245

RESULT 37
ATRX_HUMAN STANDARD; PRT; 2375 AA.
AC P46100; P51068; Q15886;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED HELICASE II) (X-LINKED
DE NUCLEAR PROTEIN) (XNP).
DE ATRX OR RAD54L OR XH2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```


AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
-!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
-!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
-!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
-!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L12104; AAA62313.1; -;
EMBL; X07273; CAA30256.1; -;
PIR; A28298; A28298.
HSSP; P08799; LMND.
InterPro; IPR000048; -;
InterPro; IPR001609; -;
InterPro; IPR002928; -;
Pfam; PF00612; IQ; 1.
Pfam; PF01576; Myosin_tail; 1.
Pfam; PF00063; myosin_head; 1.
PRINTS; PR00193; MYOSINHEAVY.
PROSITE; PS00096; IQ; 1.
Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
ATP-binding; Methylation; Alkylation; Multigene family.
DOMAIN 1 838
DOMAIN 839 1934
DOMAIN 779 801
DOMAIN 839 1934
NP_BIND 177 184
DOMAIN 654 676
DOMAIN 756 770
MOD_RES 128 128
MOD_RES 694 694
MOD_RES 704 704
CONFLICT 966 966
CONFLICT 978 978
CONFLICT 986 986
CONFLICT 1008 1014
CONFLICT 1057 1057
CONFLICT 1060 1060
CONFLICT 1095 1095
CONFLICT 1217 1217
CONFLICT 1271 1271
CONFLICT 1327 1327
CONFLICT 1358 1358
CONFLICT 1504 1504
CONFLICT 1537 1537
CONFLICT 1556 1556
SEQUENCE 1934 AA; 222928 MW; FDBAC58310B0B57D CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1934;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

557 EKIRKOLE 564
|||||||
1526 EKIRKOLE 1533

RESULT 35
MYSB_RAT
ID MYSB_RAT PRT; 1935 AA.
AC P02564;

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart.
RX MEDLINE=90016823; PubMed=2798112;
RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Complete nucleotide sequence of full length cDNA for rat beta
cardiac myosin heavy chain."
RL Nucleic Acids Res. 17:7529-7530(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=90133919; PubMed=2614840;
RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
Comparisons suggest a molecular basis for functional differences."
RL J. Mol. Biol. 210:665-671(1989).
RN [3]
RP SEQUENCE OF 1524-1935 FROM N.A.
RX MEDLINE=8220036; PubMed=7045682;
RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
RT "Molecular characterization of two myosin heavy chain genes expressed
in the adult heart."
RL Nature 297:659-664(1982).
RN [4]
RP SEQUENCE OF 1871-1935 FROM N.A.
RC STRAIN=Wistar; Tissue=Heart;
RX MEDLINE=85179510; PubMed=6241892;
RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac myosin heavy chain isozygic transitions during development
and under pathological conditions are regulated at the level of mRNA
availability."
RL Eur. Heart J. 5:181-191(1984).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- MISCELLANEOUS: THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN
THE RAT, TWO OF WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY
CHAINS.
CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CONSERVED.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X15939; CAA34065.1; -;
EMBL; J00752; AAA41654.1; -;

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AB018284; BAA34461.1; -;
EMBL; AF078035; AAD16006.1; -;
EMBL; AJ006412; CA07018.1; -;
InterPro: IPR000795; -;
Pfam; PF00009; GTP_EFTU; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding.
FT NP_BIND 638 645 GTP (BY SIMILARITY).
FT DOMAIN 39 50 POLY-LYS.
FT DOMAIN 94 99 POLY-LYS.
FT DOMAIN 138 142 POLY-LYS.
FT DOMAIN 313 322 POLY-LYS.
FT DOMAIN 353 356 POLY-LYS.
FT DOMAIN 361 364 POLY-GLU.
FT DOMAIN 491 496 POLY-GLU.
FT DOMAIN 529 567 ASP/GLU-RICH (ACIDIC).
FT MUTAGEN 640 640 V->G: LOSS OF ACTIVITY IN VIVO. RETAINS
FULL ACTIVITY IN VITRO.
FT MUTAGEN 706 706 H->E: LOSS OF ACTIVITY; BOTH IN VIVO AND
IN VITRO.
FT MUTAGEN 706 706 H->Q: LOSS OF ACTIVITY IN VIVO. PARTIAL
ACTIVITY IN VITRO.
FT MUTAGEN 759 759 D->N: LOSS OF ACTIVITY; BOTH IN VIVO AND
IN VITRO.
FT CONFLICT 64 64 G -> E (IN REF. 2).
FT CONFLICT 180 180 I -> M (IN REF. 2).
FT CONFLICT 256 256 K -> R (IN REF. 2).
FT CONFLICT 522 522 K -> T (IN REF. 2).
FT CONFLICT 549 549 E -> V (IN REF. 2).
FT CONFLICT 669 669 G -> W (IN REF. 2).
FT SEQUENCE 1220 AA; 138754 MW; 61F0435138B262E7 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1220;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 1324 PEEEEEE 1331
531 PEEEEEE 538
|||||||

RESULT 33
ID GBFL_HUMAN STANDARD; PRT; 1859 AA.
AC Q92538;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GOLGI-SPECIFIC BREFELDIN A-RESISTANCE GUANINE NUCLEOTIDE EXCHANGE
DE FACTOR 1 (BFA-RESISTANT GEF 1).
GN GBFL OR KIA0248.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047539; PubMed=9828135;
RA Mansour S.J., Herbrick J.-A., Scherer S.W., Melancon P.;
RT "Human GBFL is a ubiquitously expressed gene of the sec7 domain family
mapping to 10q24.";
RL Genomics 54:323-327(1998).
RN [2]
RP SEQUENCE OF 169-1859 FROM N.A.
RX TISSUE=Bone marrow;
MEDLINE=97191544; PubMed=9039502;

RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RT Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RA "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF5. PROMOTES
CC SIMILARITY).
CC -!- ENZYME REGULATION: INHIBITED BY BREFELDIN A (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; AF068755; AAD15903.1; -;
EMBL; D87435; BAA13379.1; -;
DR MIM; 603698; -;
DR HSSP; Q99418; 1PBV.
DR InterPro: IPR000904; -;
DR Pfam; PF01369; Sec7; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 696 884 SEC7.
FT SEQUENCE 1859 AA; 206444 MW; 5015D2BD70009CFA CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1859;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 DDEDEGVP 641
11111111
DB 1477 DDEDEGVP 1484

RESULT 34
ID MYSB_MESAU STANDARD; PRT; 1934 AA.
AC P13540; Q60540;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLB; TISSUE=Liver;
RX MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
RT heavy chain gene from Syrian hamster.";
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 962-1935 FROM N.A.
RX MEDLINE=88247788; PubMed=3380703;
RA Jandreski M.A., Sole M.J., Liew C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
RT heavy chain.";
RL Nucleic Acids Res. 16:4737-4737(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

FT	DISULFID	862	875	BY SIMILARITY.
FT	DISULFID	869	884	BY SIMILARITY.
FT	DISULFID	886	899	BY SIMILARITY.
FT	DISULFID	905	917	BY SIMILARITY.
FT	DISULFID	913	926	BY SIMILARITY.
FT	DISULFID	928	941	BY SIMILARITY.
FT	DISULFID	947	956	BY SIMILARITY.
FT	DISULFID	952	965	BY SIMILARITY.
FT	DISULFID	967	980	BY SIMILARITY.
FT	DISULFID	986	998	BY SIMILARITY.
FT	DISULFID	994	1007	BY SIMILARITY.
FT	DISULFID	1009	1023	BY SIMILARITY.
FT	DISULFID	1029	1042	BY SIMILARITY.
FT	DISULFID	1036	1051	BY SIMILARITY.
FT	DISULFID	1056	1068	BY SIMILARITY.
FT	CARBOHYD	180	180	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	507	507	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1035	1035	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	1184 AA;	126543 MW; CA48490A5F9EC5D CRC64;	

Query Match 0.6%; Score 8; DB 1; Length 1184;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1325	EEEEEE	1332	
DB	280	EEEEEE	287	

RESULT 32
IF2P_HUMAN
ID IF2P_HUMAN STANDARD; PRT; 1220 AA.
AC O60841; O95805;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSLATION INITIATION FACTOR IF-2.
GN KIAA0741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [2]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP SEQUENCE=Testis;
RC MEDLINE=99218282; PubMed=10200264;
RX Lee J.H., Choi S.K., Rolli-Mecak A., Burley S.K., Dever T.E.;
RA "Universal conservation in translation initiation revealed by human
RA and archaeal homologs of bacterial translation initiation factor
RT IF2.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
RL [3]
RN SEQUENCE OF 833-1220 FROM N.A..
RP Stanchi F., Bertocco E., Lanfranchi G., Valle G.;
RA "Finding homologues between human and yeast.";
RX Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC THE BINDING OF THE FORMYL METHIONINE-tRNA TO RIBOSOMES. SEEMS TO
CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```
FT TRANSMEM 150 166 POTENTIAL.
FT TRANSMEM 179 195 POTENTIAL.
FT TRANSMEM 212 228 POTENTIAL.
FT TRANSMEM 237 253 POTENTIAL.
FT TRANSMEM 257 273 POTENTIAL.
FT TRANSMEM 287 303 POTENTIAL.
FT DOMAIN 304 671 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 672 689 POTENTIAL.
FT TRANSMEM 700 716 POTENTIAL.
FT TRANSMEM 741 757 POTENTIAL.
FT DOMAIN 758 817 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 818 834 POTENTIAL.
FT TRANSMEM 837 853 POTENTIAL.
FT TRANSMEM 895 911 POTENTIAL.
FT DOMAIN 912 1166 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 876 876 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 80 80 K -> E (IN REF. 2).
FT CONFLICT 130 130 R -> P (IN REF. 2).
FT CONFLICT 538 538 G -> A (IN REF. 2).
FT CONFLICT 790 790 I -> L (IN REF. 2).
FT SEQUENCE 1166 AA; 130506 MW; 5042C650546E4E79 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1166;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 LLSVLVPQ 276
DB 339 LLSVLVPQ 346
|||||

RESULT 30
CYA6_HUMAN STANDARD; PRT; 1168 AA.
ID QYAG6_HUMAN STANDARD; PRT; 1168 AA.
AC Q43306; Q9NR75;
DE 15-JUL-1999 (Rel. 38, Created)
DE 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE (CA(2+)-INHIBITABLE ADENYLYL CYCLASE).
DE ADCY6 OR KIAA0422.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Thyroid;
Wicker R., Gascon Catalan A., Cailleux A.-F., Starenki D., Stengel D.,
Sarasin A., Suarez H.G.;
"Cloning and expression of human adenylyl cyclase type VI from normal
thyroid tissues.";
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 142-1168 FROM N.A.
TISSUE=Brain;
MEDLINE=98116655; PubMed=9455477;
Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
"Prediction of the coding sequences of unidentified human genes. VIII.
78 new cDNA clones from brain which code for large proteins in
vitro.";
DNA Res. 4:307-313(1997).
CC -!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL
CYCLASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR
CONCENTRATION RANGE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
```

```
CC CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF250226; AAF82478.1; -.
CC EMBL; AB007882; BAA24852.1; -.
CC MIM; 600294; -.
CC
CC Pfam; PF00211; guanylate_cyc; 2.
CC PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
CC PROSITE; PS0125; GUANYLATE_CYCLASES_2; 2.
CC Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
CC DOMAIN 1 151 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 152 168 POTENTIAL.
CC TRANSMEM 181 197 POTENTIAL.
CC TRANSMEM 214 230 POTENTIAL.
CC TRANSMEM 239 255 POTENTIAL.
CC TRANSMEM 259 275 POTENTIAL.
CC TRANSMEM 289 305 POTENTIAL.
CC DOMAIN 306 673 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 674 691 POTENTIAL.
CC TRANSMEM 702 718 POTENTIAL.
CC TRANSMEM 743 759 POTENTIAL.
CC DOMAIN 760 819 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 820 836 POTENTIAL.
CC TRANSMEM 839 855 POTENTIAL.
CC TRANSMEM 897 913 POTENTIAL.
CC DOMAIN 914 1168 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 793 793 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 762 814 MISSING (IN REF. 2).
CC SEQUENCE 1168 AA; 130615 MW; 127BB6E67F73AA61 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1168;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 LLSVLVPQ 276
DB 341 LLSVLVPQ 348
|||||

RESULT 31
FBL2_HUMAN STANDARD; PRT; 1184 AA.
ID FBL2_HUMAN STANDARD; PRT; 1184 AA.
AC P98095;
DE 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE FIBULIN-2 PRECURSOR.
GN FBLN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Fibroblast;
RX MEDLINE=95104855; PubMed=7806230;
RA Zhang R.-Z., Pan T.-C., Zhang Z.-Y., Mattei M.-G., Timpl R.,
Chu M.-L.;
"Fibulin-2 (FBLN2): human cDNA sequence, mRNA expression, and mapping
of the gene on human and mouse chromosomes.";
RL Genomics 22:425-430(1994).
CC -!- FUNCTION: ITS BINDING TO FIBRONECTIN AND SOME OTHER LIGANDS IS
CC CALCIUM DEPENDENT.
CC -!- SUBUNIT: DISULFIDE-LINKED HOMOTRIMER.
```


(CA(2+)-INHIBITABLE ADENYL CYCLASE).
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
MEDLINE=92357702; PubMed=1379717;
Yoshimura M., Cooper D.M.F.;
"Cloning and expression of a Ca(2+)-inhibitable adenylyl cyclase from
NCB-20 cells";
Proc. Natl. Acad. Sci. U.S.A. 89:6716-6720(1992).
[2]
SEQUENCE OF 10-1165 FROM N.A.
MEDLINE=93076707; PubMed=1332848;
Premont R.T., Jacobowitz O., Ivengar R.;
"Lowered responsiveness of the catalyst of adenylyl cyclase to
stimulation by GS in heterologous desensitization: a role for
adenosine 3',5'-monophosphate-dependent phosphorylation";
Endocrinology 131:2774-2784(1992).
-!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYL
CYCLASE.
-!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
-!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR
CONCENTRATION RANGE.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- TISSUE SPECIFICITY: MOST ABUNDANT IN HEART BUT WEAKLY DETECTABLE
IN BRAIN, INTESTINE, LUNG, AND SPLEEN.
-!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
-!- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-4/GUANYLYL CYCLASE
FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M93422; AAA37174.1; -.
EMBL; M96653; AAA37182.1; -.
PIR; A46187; A46187.
HSP; P19754; IAWK.
MGD; MGI:87917; Adcy6.
InterPro; IPR001054; -.
Pfam; PF00211; guanylate_cyc; 2.
PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
PROSITE; PS00125; GUANYLATE_CYCLASES_2; 2.
Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
DOMAIN 1 149
CYTOPLASMIC (POTENTIAL).
TRANSMEM 150 166
TRANSMEM 179 195
TRANSMEM 212 228
TRANSMEM 237 253
TRANSMEM 257 273
TRANSMEM 287 303
DOMAIN 304 670
CYTOPLASMIC (POTENTIAL).
TRANSMEM 671 688
TRANSMEM 699 715
TRANSMEM 740 756
DOMAIN 757 816
TRANSMEM 817 833
TRANSMEM 836 852
TRANSMEM 894 910
DOMAIN 911 1165
CYTOPLASMIC (POTENTIAL).
CARBOHYD 277 277
CARBOHYD 790 790
CARBOHYD 875 875
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
K -> N (IN REF. 2).
CONFLICT 47 47
CONFLICT 76 76
GR -> RAG (IN REF. 2).
CONFLICT 508 509

FT CONFLICT 737 737 V -> G (IN REF. 2).
FT CONFLICT 881 881 L -> Q (IN REF. 2).
FT CONFLICT 990 990 V -> M (IN REF. 2).
SQ SEQUENCE 1165 AA; 130318 MW; 24EE1BB45DF1E87E CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1165;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 LLSVLPQ 276
Db 339 LLSVLPQ 346

RESULT 29
CYA6_RAT
ID CYA6_RAT STANDARD; PRT; 1166 AA.
AC Q03343;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE (CA(2+)-INHIBITABLE ADENYL CYCLASE).
GN ADCY6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93077589; PubMed=1332969;
RA Krupinski J., Lehman T.C., Frankenfield C.D., Zwaagstra J.C.,
Watson P.A.;
RT "Molecular diversity in the adenylyl cyclase family. Evidence for
eight forms of the enzyme and cloning of type VI";
RL J. Biol. Chem. 267:24858-24862(1992).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93028552; PubMed=1409703;
RA Premont R.T., Chen J., Ma H.-W., Ponnappalli M., Ivengar R.;
RT "Two members of a widely expressed subfamily of hormone-stimulated
adenylyl cyclases";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9809-9813(1992).
CC -!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYL
CYCLASE.
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR
CONCENTRATION RANGE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-4/GUANYLYL CYCLASE
FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L01115; AAA40676.1; -.
EMBL; M96160; AAA40678.1; ALT_INIT.
PIR; A45145; A45145.
HSP; P19754; IAWK.
InterPro; IPR001054; -.
Pfam; PF00211; guanylate_cyc; 2.
PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
PROSITE; PS00125; GUANYLATE_CYCLASES_2; 2.
Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
DOMAIN 1 149
CYTOPLASMIC (POTENTIAL).
TRANSMEM 150 166
TRANSMEM 179 195
TRANSMEM 212 228
TRANSMEM 237 253
TRANSMEM 257 273
TRANSMEM 287 303
DOMAIN 304 670
CYTOPLASMIC (POTENTIAL).
TRANSMEM 671 688
TRANSMEM 699 715
TRANSMEM 740 756
DOMAIN 757 816
TRANSMEM 817 833
TRANSMEM 836 852
TRANSMEM 894 910
DOMAIN 911 1165
CYTOPLASMIC (POTENTIAL).
CARBOHYD 277 277
CARBOHYD 790 790
CARBOHYD 875 875
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
K -> N (IN REF. 2).
CONFLICT 47 47
CONFLICT 76 76
GR -> RAG (IN REF. 2).
CONFLICT 508 509

15-DEC-1998 (Rel. 37, Last annotation update)
SUCROSE-PHOSPHATE SYNTHASE 2 (EC 2.4.1.14) (UDP-GLUCOSE-FRUCTOSE-
PHOSPHATE GLUCOSYLTRANSFERASE 2).
SPS2.
Craterostigma plantaginum.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
Lamiales; Scrophulariaceae; Craterostigma.
NCBI_TaxID=4153;
[1]
SEQUENCE FROM N.A.
MEDLINE=97451773; PubMed=9306694;
Ingram J., Chandler J.W., Gallagher L., Salamini F., Bartels D.;
"Analysis of cDNA clones encoding sucrose-phosphate synthase in
relation to sugar interconversions associated with dehydration in the
resurrection plant Craterostigma plantaginum Hochst.";
Plant Physiol. 113:113-121(1997).
-!- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
PHOTOASSIMILATES OUT OF THE LEAF.
-!- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE 6-PHOSPHATE = UDP +
SUCROSE 6-PHOSPHATE.
-!- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
-!- PATHWAY: SUCROSE SYNTHESIS PATHWAY.
-!- SUBUNIT: HOMODIMER OR HOMOTETRAMER (BY SIMILARITY).
-!- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
ENZYME FUNCTION (BY SIMILARITY).
-!- SIMILARITY: SOME, WITH SUCROSE SYNTHASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: Y11795; CAA72491.1; .
InterPro: IPR001296; .
Pfam: PF00534; Glycosyltransferase; Phosphorylation; Multigene family.
Transferrase; Glycosyltransferase; POLY-SER.
DOMAIN 245 248
DOMAIN 256 264 POLY-GLU.
DOMAIN 787 790 POLY-ARG.
SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;
Query Match 0.6%; Score 8; DB 1; Length 1081;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
|||||||
DB 257 PEEEEEE 264
RESULT 27
CYA6_CANFA STANDARD; PRT; 1165 AA.
AC P30804;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE (CA(2+))-INHIBITABLE ADENYLYL CYCLASE).
DE ADCY6.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
[1]
SEQUENCE FROM N.A.
TISSUE=Heart;
MEDLINE=92409599; PubMed=1528892;
Katsushika S., Chen L., Kawabe J., Nilakantan R., Halnon N.J.,
Honey C.J., Ishikawa Y.;
"Cloning and characterization of a sixth adenylyl cyclase isoform:
types V and VI constitute a subgroup within the mammalian adenylyl
cyclase family.";
Proc. Natl. Acad. Sci. U.S.A. 89:8774-8778(1992).
-!- FUNCTION: THIS A MEMBRANE-BOUND, CA(2+)-INHIBITABLE ADENYLYL
CYCLASE.
-!- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.
-!- ENZYME REGULATION: INHIBITION BY CA(2+) IN THE SUBMICROMOLAR
CONCENTRATION RANGE.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
-!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: M94968; .; NOT_ANNOTATED_CDS.
PIR: A46180; A46180.
HSP: P19754; IAWK.
InterPro: IPR001054; .
Pfam: PF00211; guanylate_cyc; 2.
PROSITE: P500452; GUANYLATE_CYCLASES_1; 2.
PROSITE: P50125; GUANYLATE_CYCLASES_2; 2.
Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
DOMAIN 1 149
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 150 166
POTENTIAL.
FT TRANSMEM 179 195
POTENTIAL.
FT TRANSMEM 212 228
POTENTIAL.
FT TRANSMEM 237 253
POTENTIAL.
FT TRANSMEM 257 273
POTENTIAL.
FT TRANSMEM 287 303
POTENTIAL.
FT DOMAIN 304 670
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 671 688
POTENTIAL.
FT TRANSMEM 699 715
POTENTIAL.
FT TRANSMEM 740 756
POTENTIAL.
FT DOMAIN 757 816
EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 817 833
POTENTIAL.
FT TRANSMEM 836 852
POTENTIAL.
FT TRANSMEM 894 910
POTENTIAL.
FT DOMAIN 911 1165
CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 790 790
N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875
N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1165 AA; 130323 MW; BA9D2D329120615E CRC64;
Query Match 0.6%; Score 8; DB 1; Length 1165;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 269 LLLSVLPQ 276
|||||||
DB 339 LLLSVLPQ 346
RESULT 28
CYA6_MOUSE
ID CYA6_MOUSE STANDARD; PRT; 1165 AA.
AC Q01341;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENYLATE CYCLASE, TYPE VI (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE

RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordisiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehman H., Reinhardt R., Yaspo M.-L.,
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -!- FUNCTION: MAY DEUBIQUITINATE ONE OR MORE CRITICAL PROTEINS THAT
CC ARE INVOLVED IN THE CONDENSATION OF MITOTIC CHROMOSOMES, POSSIBLY
CC ACTING SELECTIVELY ON HISTONES H2A AND H2B, THE MAJOR
CC UBIQUITINATED PROTEINS OF CHROMATIN. IT IS ABLE TO DEUBIQUITINATE
CC HISTONE H2A IN VITRO, THE PHOSPHORYLATED FORM OF THE PROTEIN IS
CC ALSO ENZYMATICALLY ACTIVE.
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
CC UBIQUITIN + A THIOL.
CC -!- TISSUE SPECIFICITY: PRESENT IN ALL THE TISSUES EXAMINED INCLUDING
CC FETAL BRAIN, LUNG, LIVER, KIDNEY, AND ADULT HEART, BRAIN,
CC PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
CC -!- PTM: PHOSPHORYLATED AT THE ONSET OF MITOSIS AND
CC DEPHOSPHORYLATED DURING THE METAPHASE/ANAPHASE TRANSITION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF126736; AAD20949.1; -;
CC EMBL; AL163249; CAB90432.1; -;
CC MEROPS; C19.021; -;
CC MIM; 604735; -;
CC InterPro; IPR001394; -;
CC Pfam; PF00442; UCH-1; 1.
CC Pfam; PF00443; UCH-2; 1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS0235; UCH_2_3; 1.
CC Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family;
KW Phosphorylation.
FT ACT_SITE 205 205 BY SIMILARITY.
FT ACT_SITE 750 750 BY SIMILARITY.
FT ACT_SITE 758 758 BY SIMILARITY.
FT MUTAGEN 205 205 C->S: UNABLE TO DEUBIQUITINATE H2A/H2B.
SQ SEQUENCE 823 AA; 93570 MW; C7D4175649BA3E31 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 823;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 DODSEEEK 53
DQ 412 DODSEEEK 419

RESULT 25
IF2P_YEAST STANDARD; PRT; 1002 AA.
AC P39730;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSLATION INITIATION FACTOR IF-2.

GN YAL035W OR FUN12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94357438; PubMed=8076820;
RA Sutcliffe P., Shafer B.K., Strathern J.N., Hughes S.H.;
RT "Isolation, identification and characterization of the FUN12 gene of
RT Saccharomyces cerevisiae.";
RL Gene 146:209-213(1994).
RN [3]
RP REVISIONS
RC STRAIN=S288C / AB972;
RA Vo D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION
RX MEDLINE=98289357; PubMed=9624054;
RA Choi S.K., Lee J.H., Zoll W.L., Merrick W.C., Dever T.E.;
RT "Promotion of met-tRNA^{Met} binding to ribosomes by yif2, a bacterial
RT IF2 homolog in yeast.";
RL Science 280:1757-1760(1998).
CC -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO
CC FUNCTION ALONG WITH EIF-2.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12980; AAC04996.1; -;
CC EMBL; L29389; AAA57228.1; ALT_SEQ.
CC SGD; S0000033; FUN12.
CC InterPro; IPR000795; -;
CC Pfam; PF00009; GTP_EFTU; 1.
CC Initiation factor; Protein biosynthesis; GTP-binding.
KW DOMAIN 361 371 POLY-GLU.
FT NP_BIND 412 419 GTP (BY SIMILARITY).
SQ SEQUENCE 1002 AA; 112268 MW; 1A496195DAE1C283 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 1002;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DQ 365 EEEEEER 372

RESULT 26
SPS2_CRAPL STANDARD; PRT; 1081 AA.
AC 004933;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

-!- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS FARNESYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE ENVELOPE INTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.

-!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.

-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

THIS IS A B TYPE LAMIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X16705; CAA34677.1; -
EMBL: M35153; AAC96023.1; ALT_SEQ.
PIR: S07720; S07720.
SWISS-2DPAGE: P14733; MOUSE.
MGD: MGI:96795; Lmmb1.
InterPro: IPR001322; -
InterPro: IPR001664; -
Pfam: PF00932; IF_tail; 1.
Pfam: PF00038; filament; 1.
PROSITE: PS00226; IF; 1.
Nuclear protein; Lipoprotein; Prenylation; Phosphorylation.
INIT_MET 0
DOMAIN 1 34 HEAD.
DOMAIN 35 386 ROD.
DOMAIN 387 587 TAIL.
DOMAIN 35 69 COIL 1A.
DOMAIN 70 81 LINKER 1.
DOMAIN 82 215 COIL 1B.
DOMAIN 216 243 LINKER 2.
DOMAIN 244 386 COIL 2.
DOMAIN 415 420 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
DOMAIN 552 560 GLU-RICH (HIGHLY ACIDIC, COULD BE INVOLVED IN CHROMATIN BINDING).
LIPID 584 584 FARNESYL (BY SIMILARITY).
SEQUENCE 587 AA; 66753 MW; 222E87CDF053E4B7 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
DB 551 PEEEEEE 558
|||||

RESULT 23
APPL_MOUSE STANDARD; PRT; 653 AA.
AC Q03157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
GN APLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93066322; PubMed=1279693;

RA Wasco W., Bupp K., Magendantz M., Gusella J.F., Tanzi R.E., Solomon F.;
RT "Identification of a mouse brain cDNA that encodes a protein related to the Alzheimer disease-associated amyloid beta protein precursor."; Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP IS SECRETED.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: L04538; AAA37247.1; -
DR PIR: A46362; A46362.
DR MGD: MGI:88046; Aplp1.
DR InterPro: IPR001868; -
DR PRINTS: PR00203; AMYLOIDA4.
DR PROSITE: PS00319; A4_EXTRA; 1.
DR PROSITE: PS00320; A4_INTRA; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 653 AMYLOID-LIKE PROTEIN 1.
FT DOMAIN 21 583 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 584 606 POTENTIAL.
FT DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 643 646 CLATHRIN-BINDING (POTENTIAL).
FT DOMAIN 263 271 POLY-GLU.
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 653 AA; 72751 MW; 56516DC3EA40E4B0 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 653;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332
DB 265 EEEEEEE 272
|||||

RESULT 24
UBPG_HUMAN STANDARD; PRT; 823 AA.
AC Q9Y3T5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 16 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 16) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 16)
DE (DEUBIQUITINATING ENZYME 16) (UBIQUITIN PROCESSING PROTEASE UBP-M).
GN USP16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99178974; PubMed=10077596;
RA Cai S.-Y., Babbitt R.W., Marchesi V.T.;
RT "A mutant deubiquitinating enzyme (Ubp-M) associates with mitotic chromosomes and blocks cell division."; Proc. Natl. Acad. Sci. U.S.A. 96:2828-2833(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,

Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1323 APEEEEE 1330

|||||

462 APEEEEE 469

RESULT 21

LAM1_HUMAN

AC P20700; STANDARD; PRT; 585 AA.

01-FEB-1991 (Rel. 17, Created)

01-FEB-1991 (Rel. 17, Last sequence update)

30-MAY-2000 (Rel. 39, Last annotation update)

LAMIN B1.

LMNB1 OR LMN2 OR LMNB.

Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

MEDLINE-90220602; PubMed-2325650;

Pollard K.M., Chan E.K.B., Grant B.J., Sullivan K.F., Tan E.M.,

Glass C.A.;

"In vitro posttranslational modification of lamin B cloned from a

human T-cell line.";

Mol. Cell. Biol. 10:2164-2175(1990).

[2]

SEQUENCE FROM N.A.

MEDLINE-96044426; PubMed-7557986;

Lin F., Worman H.J.;

"Structural organization of the human gene (LMNB1) encoding nuclear

lamin B1.";

Genomics 27:230-236(1995).

[3]

ISOPRENOID.

MEDLINE-90062174; PubMed-2684976;

Farnsworth C.C., Wolda S.L., Gelb M.H., Glomset J.A.;

"Human lamin B contains a farnesylated cysteine residue.";

J. Biol. Chem. 264:20422-20429(1989).

-!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS

LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,

WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE

AND MAY ALSO INTERACT WITH CHROMATIN.

-!- SUBCELLULAR LOCATION: NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR

MEMBRANE.

-!- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS

FARNESYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF

THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY

PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.

-!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY

CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND

FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,

RESPECTIVELY.

-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

THIS IS A B TYPE LAMIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M34458; AAC36162.1; -

EMBL; L37747; AAC37575.1; -

EMBL; L37737; AAC37575.1; JOINED.

EMBL; L37738; AAC37575.1; JOINED.

EMBL; L37739; AAC37575.1; JOINED.

EMBL; L37740; AAC37575.1; JOINED.

DR

DR EMBL; L37741; AAC37575.1; JOINED.
DR EMBL; L37742; AAC37575.1; JOINED.
DR EMBL; L37743; AAC37575.1; JOINED.
DR EMBL; L37744; AAC37575.1; JOINED.
DR EMBL; L37745; AAC37575.1; JOINED.
DR EMBL; L37746; AAC37575.1; JOINED.
DR PIR; A34707; A34707.
DR SWISS-2DPAGE; P20700; HUMAN.
DR Aarhus/Ghent-2DPAGE; 7510; IEF.
DR MIM; 150340; -
DR InterPro; IPR001322; -
DR InterPro; IPR001664; -
DR Pfam; PF00932; IF_tail; 1.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil;
KW Nuclear protein; Lipoprotein; Prenylation; Phosphorylation.
FT INIT_MET 0
FT DOMAIN 1 33 HEAD.
FT DOMAIN 34 385 ROD.
FT DOMAIN 386 585 TAIL.
FT DOMAIN 34 68 COIL 1A.
FT DOMAIN 69 80 LINKER 1.
FT DOMAIN 81 214 COIL 1B.
FT DOMAIN 215 242 LINKER 2.
FT DOMAIN 243 385 COIL 2.
FT DOMAIN 414 419
FT DOMAIN 551 558
FT LIPID 582 582 FARNESYL.
SQ SEQUENCE 585 AA; 66277 MW; 0802EDAC33CBF998 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 585;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

|||||

Db 550 PEEEEEE 557

RESULT 22

LAM1_MOUSE

ID LAM1_MOUSE STANDARD; PRT; 587 AA.

AC P14733;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE LAMIN B1.

GN LMNB1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RC SEQUENCE FROM N.A.

RP TISSUE=Liver;

RX MEDLINE=89210899; PubMed=3243285;

RA Hoeger T.H., Krohne G., Franke W.W.;

"Amino acid sequence and molecular characterization of murine lamin B

as deduced from cDNA clones.";

RL Eur. J. Cell Biol. 47:283-290(1989).

[2]

RP REVISIONS.

RA Hoeger T.H.;

RL Submitted (OCT-1989) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS

LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,

WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE

AND MAY ALSO INTERACT WITH CHROMATIN.

CC -!- SUBCELLULAR LOCATION: NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR

MEMBRANE.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U70158; AAC50135.1; -
MIM: 601603; -
HSSP: P15054; LBKM.
InterPro: IPR000980; -
Pfam: PF00017; SH2; 1.
PROSITE: PS50001; SH2; 1.
SH2 domain; Phosphorylation. POLY-GLU.
DOMAIN 133 136
FT DOMAIN 198 201 POLY-PRO.
FT DOMAIN 422 530 SH2.
FT SEQUENCE 533 AA; 60188 MW; C5D22F31D36200C8 CRC64;
SQ

Query Match 0.6%; Score 8; DB 1; Length 533;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 SSPLPGKK 1197
|||||
DB 297 SSPLPGKK 304

RESULT 19
INVO_AOTTR STANDARD; PRT; 544 AA.
AC P24708;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INVOLUCRIN.
GN IVL.
OS Aotus trivirgatus (Night monkey) (Douroucoulis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX NCBI_TaxID=9505;
RN [1]
RE SEQUENCE FROM N.A.
RX MEDLINE=90014142; PubMed=2507864;
RA Tseng H., Green H.;
RT "The involucrin gene of the owl monkey: origin of the early region.";
RI Mol. Biol. Evol. 6:460-468(1989).
CC -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M25313; AAA5375.1; -
HSSP: P80220; 1DIP.
InterPro: IPR000354; -
Pfam: PF00094; Involutrin; 42.
PROSITE: PS00795; INVOLUCRIN; 1.
KW Keratinocyte; Repeat.
SQ SEQUENCE 544 AA; 63927 MW; 2A02ABA5E1499F9D CRC64;
SQ

Query Match 0.6%; Score 8; DB 1; Length 544;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 SQELLDTV 199
|||||
DB 16 SQELLDTV 23

RESULT 20
NFL_XENLA STANDARD; PRT; 544 AA.
AC P35616;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET L PROTEIN (NEUROFILAMENT LIGHT POLYPEPTIDE) (NF-L).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RE SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=92356194; PubMed=1494944;
RA Charnas L.R., Szaro B.G., Gainer H.;
RT "Identification and developmental expression of a novel low molecular weight neuronal intermediate filament protein expressed in Xenopus laevis.";
RI J. Neurosci. 12:3010-3024(1992).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M, AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 10-NM FILAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M86654; AAA83018.1; -
PIR: B44841; B44841.
InterPro: IPR001684; -
Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone.
FT DOMAIN 1 87 HEAD.
FT DOMAIN 88 390 ROD.
FT DOMAIN 391 544 TAIL.
FT DOMAIN 88 119 COIL 1A.
FT DOMAIN 120 132 LINKER 1.
FT DOMAIN 133 228 COIL 1B.
FT DOMAIN 229 246 LINKER 12.
FT DOMAIN 247 265 COIL 2A.
FT DOMAIN 266 274 LINKER 2.
FT DOMAIN 275 390 COIL 2B.
FT DOMAIN 391 435 TAIL, SUBDOMAIN A.
FT DOMAIN 436 544 TAIL, SUBDOMAIN B (ACIDIC).
FT DOMAIN 441 538 GLU-RICH.
FT DOMAIN 464 469 POLY-GLU.
SQ SEQUENCE 544 AA; 61861 MW; 76D911B896E97201 CRC64;
SQ

Query Match 0.6%; Score 8; DB 1; Length 544;

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X06832; CAA29988.1; --
PIR; A23996; A23996.
PIR; S0291; S0291.
PIR; S02543; S02543.
InterPro; IPR001819; --
InterPro; IPR001990; --
Pfam; PF01271; Granin; 1.
PRINTS; PR00659; CHROMOGRAIN.
PROSITE; PS00422; GRANINS_1; 1.
PROSITE; PS00423; GRANINS_2; 1.
Signal; Amidation; Glycoprotein; Calcium-binding.
SIGNAL 18
CHAIN 19 466
BETA-GRANIN
PEPTIDE 19 146
PEPTIDE 281 332
PEPTIDE 361 374
DOMAIN 92 110
DOMAIN 231 249
DOMAIN 344 350
MOD_RES 332 332
DISULFID 35 56
SEQUENCE 466 AA; 52024 MW; 05D135FFA657C48C CRC64;

Query Match 0.6%; Score 8; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DR 344 EEEEEER 351

RESULT 17

NIFK_RHISN STANDARD; PRT; 513 AA.
AC P19067;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NITROGENASE MOLYBDENUM-IRON PROTEIN BETA CHAIN (EC 1.18.6.1)
DE (NITROGENASE COMPONENT I) (DINITROGENASE).
GN (NIFK1 OR Y4VM) AND (NIFK2 OR Y4XC).
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Feilley R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
RN [2]
RP SEQUENCE OF 132-195 FROM N.A.
RC STRAIN=ANU 240;
RX MEDLINE=89306671; PubMed=2744485;
RA Badenoch-Jones J., Holton T.A., Morrison C.M., Scott K.F., Shine J.;
RT "Structural and functional analysis of nitrogenase genes from the
broad-host-range Rhizobium strain ANU240.";
RL Gene 77:141-153(1989).
CC -!- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE

CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
CC -!- CATALYTIC ACTIVITY: 8 REDUCED FERREDOXIN + 8 H(+) + N(2) + 16 ATP
CC = 8 OXIDIZED FERREDOXIN + 2 NH(3) + 16 ADP + 16 ORTHOPHOSPHATE.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains that binds
CC 30-32 Fe, 2 Mo, and inorganic sulfur.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; M26963; AAA26327.1; --
EMBL; AE000102; AAB91901.1; --
EMBL; AE000105; AAB91925.1; --
PIR; PS0046; PS0046.
HSSP; P07329; IN2C.
InterPro; IPR000318; --
InterPro; IPR000510; --
Pfam; PF00148; Oxidored_nitro; 1.
PROSITE; PS00699; NITROGENASE_1; 1.
PROSITE; PS00090; NITROGENASE_1_2; 1.
KW Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur; Plasmid;
KW Multigene family.
SQ SEQUENCE 513 AA; 57302 MW; 41631040335541AE CRC64;

Query Match 0.6%; Score 8; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 LRLLTTL 858
DB 481 LRLLTTL 488

RESULT 18

LCP2_HUMAN STANDARD; PRT; 533 AA.
AC Q13094;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LYMPHOCYTE CYTOSOLIC PROTEIN 2 (SH2 DOMAIN-CONTAINING LEUCOCYTE
DE PROTEIN OF 76 KDA) (SLP-76 TYROSINE PHOSPHOPROTEIN) (SLP76).
GN LCP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Leukemia;
RX MEDLINE=95221345; PubMed=7706237;
RA Jackman J.K., Motto D.G., Sun Q., Tanemoto M., Turck C.W., Peltz G.A.,
RA Koretzky G.A., Findell P.R.;
RT "Molecular cloning of SLP-76, a 76-kDa tyrosine phosphoprotein
associated with Grb2 in T cells.";
RL J. Biol. Chem. 270:7029-7032(1995).
CC -!- FUNCTION: INVOLVED IN T CELL ANTIGEN RECEPTOR MEDIATED SIGNALING.
CC -!- SUBUNIT: INTERACTS WITH THE ADAPTER PROTEINS GRB2 AND Fyb.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN, THYMUS, AND
CC PERIPHERAL BLOOD LEUKOCYTES. HIGHLY EXPRESSED ALSO IN T CELL AND
CC MONOCYTIC CELL LINES. EXPRESSED AT LOWER LEVEL IN B CELL LINES.
CC NOT DETECTED IN FIBROBLAST OR NEUROBLASTOMA CELL LINES.
CC -!- PTM: PHOSPHORYLATED AFTER T-CELL RECEPTOR ACTIVATION.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -----

EMBL; X07547; CAA30421.1; -
DR EMBL; X06707; CAA29892.1; -
DR EMBL; J03321; AAA91569.1; -
DR EMBL; M19487; AAB02586.1; ALT_FRAME.
DR EMBL; Y00505; CAA68558.1; ALT_FRAME.
DR PIR; S01921; S01921.
DR PIR; S01022; S01022.
DR PIR; S01023; S01023.
DR PIR; S06324; S06324.
DR InterPro; IPR001198; -
DR Pfam; PF00772; DnaB; 1.
KW DNA replication; DNA-binding; ATP-binding; Helicase; Primosome;
Plasmid.
FT NP_BIND 225 232 ATP (POTENTIAL).
FT VARIANT 55 55 Y -> H (IN PLASMID PCHL1).
FT VARIANT 188 188 A -> V (IN PLASMID PCHL1).
FT VARIANT 189 189 T -> A (IN PLASMID PCHL1).
FT VARIANT 200 200 T -> A (IN PLASMID PCHL1 AND PCTT1).
FT VARIANT 200 200 A -> E (IN PLASMID PCHL1).
FT VARIANT 200 200 A -> G (IN PLASMID PCTT1).
FT VARIANT 267 267 I -> V (IN PLASMID PCTT1).
SQ SEQUENCE 451 AA; 51456 MW; 2CD73A152C72A1F1 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 451;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1169 RIKGRDLS 1176
DB 145 RIKGRDLS 152

RESULT 15
CMGA_MOUSE STANDARD; PRT; 463 AA.
AC P26339;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CHROMOGANIN A PRECURSOR (CGA) [CONTAINS: PANCREASTATIN; BETA-GRANIN;
WE-14].
GN CHGA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91302337; PubMed=2071596;
RA Wu H.J., Rozansky D.J., Farmer R.J., Gill B.M., O'Connor D.T.;
RT "Structure and function of the chromogranin A gene. Clues to
evolution and tissue-specific expression."
RL J. Biol. Chem. 266:13130-13134(1991).
CC -!- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN
RELEASE FROM THE PANCREAS.
CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
GRANULES.
CC -!- PTM: CGA IS O-GLYCOSYLATED.
CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC -!- SIMILARITY: BELONGS TO THE CHROMOGANIN / SECRETOGANIN PROTEIN
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M64278; AAA37457.1; -
DR PIR; A39868; A39868.
DR MGI; MGI:88394; Chga.

DR InterPro; IPR001819; -
DR InterPro; IPR001990; -
DR Pfam; PF01271; Granin; 1.
DR PRINTS; PR00659; CHROMOGANIN.
DR PROSITE; PS00422; GRANINS_1; 1.
DR PROSITE; PS00423; GRANINS_2; 1.
KW Signal; Amidation; Glycoprotein; Calcium-binding.
FT SIGNAL 1 18
FT CHAIN 19 463 CHROMOGANIN A.
FT PEPTIDE 19 151 BETA-GRANIN (BY SIMILARITY).
FT PEPTIDE 276 329 PANCREASTATIN (BY SIMILARITY).
FT PEPTIDE 358 371 WE-14 (BY SIMILARITY).
FT DOMAIN 103 116 POLY-GLN.
FT DOMAIN 236 245 POLY-GLU.
FT DOMAIN 340 347 POLY-GLU.
FT MOD_RES 329 329 AMIDATION (PROBABLE).
FT DISULFID 35 56 BY SIMILARITY.
SQ SEQUENCE 463 AA; 51789 MW; 1AB3C5FF433C39E4 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 463;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
DB 341 EEEEEER 348

RESULT 16
CMGA_RAT STANDARD; PRT; 466 AA.
AC P10354;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CHROMOGANIN A PRECURSOR (CGA) [CONTAINS: PANCREASTATIN; BETA-GRANIN;
WE-14].
GN CHGA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88112232; PubMed=2828116;
RA Iacangelo A., Okayama H., Eiden L.E.;
RT "Primary structure of rat chromogranin A and distribution of its
mRNA."
RL FEBS Lett. 227:115-121(1988).
RN [2]
RP SEQUENCE OF 29-466 FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=88312980; PubMed=3044825;
RA Huton J.C., Nielsen E., Kasten W.;
RT "The molecular cloning of the chromogranin A-like precursor of beta-
granin and pancreastatin from the endocrine pancreas."
RT FEBS Lett. 236:269-274(1988).
RN [3]
RP SEQUENCE OF 19-32.
RX MEDLINE=85285598; PubMed=3896848;
RA Huton J.C., Hansen F., Peshavaria M.;
RT "Beta-granins: 21 kDa co-secreted peptides of the insulin granule
closely related to adrenal medullary chromogranin A."
RL FEBS Lett. 188:336-340(1985).
CC -!- FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN
RELEASE FROM THE PANCREAS.
CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
GRANULES.
CC -!- PTM: CGA IS O-GLYCOSYLATED.
CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC -!- SIMILARITY: BELONGS TO THE CHROMOGANIN / SECRETOGANIN PROTEIN
FAMILY.
CC


```

QY      1324  PEEEEEE 1331
          |||||
Db      369  PEEEEEE 376

RESULT  14
GPID CHLTR

```

ID	GP1D_CHLTR	STANDARD;	PRT;	451 AA.
AC	P10555;	P08781;	P22445;	
DT	01-NOV-1988	(Rel. 09, Created)		
DT	01-JUL-1989	(Rel. 11, Last sequence update)		
DT	01-OCT-2000	(Rel. 40, Last annotation update)		
DE	PROBABLE PLASMID REPLICATIVE DNA HELICASE (DNAB-LIKE PROTEIN)			
DE	(VIRULENCE PLASMID PROTEIN PGPI-D) (PROTEIN P-3).			

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L2/434/BU; PLASMID=plGV440;
RX MEDLINE=89013895; PubMed=2845228;
RA Comanducci M, Ricci S, Ratti G;
RB

KL required for growth within mammalian cells.-;
 RL Mol. Microbiol. 2:531-538 (1988).
 [2]
 RN SEQUENCE FROM N.A.
 RN STRAIN=L1/440/LN; PLASMID=pLGV440;
 RC MEDLINE=88233998; PubMed=2836808;
 RX Hatt C., Ward N.E., Clarke I.N.;
 RA

RT replication.",
RL Nucleic Acids Res. 16:4053-4067(1988).
RN [3]
RP SEQUENCE FROM N. A.
RC STRAIN-G0/86 / SEROTYPE D: PLASMID=pCHL1;
RX MEDLINE=90301736; Pubmed=2194229;
RA Comanducci M., Ricci S., Cevenini R., Ratti G.;

RL plasmid 23:149-154(1990).
RN [4]
RC SEQUENCE FROM N.A.
RP STRAIN=SEROTYPE B; PLASMID=pcTTL;
RX MEDLINE=88177106; PubMed=344859;
RT Sripirakash K.S., Macavoy E.S.;
RA "Characterization and sequence of a plasmid from the trachoma biovar

[5]
RN
RP
RC STRAIN=SEROTYPE B; PLASMID=pCTT1;
RX MEDLINE=88096599; PubMed=2827121;
RA Sriprakash K.S., Macavoy E.S.;
RT "A gene for dnaB like protein in chlamydial plasmid.";
RL Nucleic Acids Res. 15:10596-10596(1987).

--

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute where are no restrictions on its use.

--

- I- SIMILARITY: BELONGS TO THE DNAB SUBFAMILY OF HELICASES.
CC
CC
CC - I- CAUTION: REF.4 AND 5 SEQUENCES DIFFER FROM THAT SHOWN DUE TO A
CC
CC FRAMESHIFT IN POSITION 389.
CC

--

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC

Wilber J.F.;
"Cloning and structure of human genomic DNA and hypothalamic cDNA
encoding human prepro thyrotropin-releasing hormone.";
Mol. Endocrinol. 4:551-556(1990).
-!- FUNCTION: TRH FUNCTIONS AS A REGULATOR OF THE BIOSYNTHESIS OF TSH
IN THE ANTERIOR PITUITARY GLAND AND AS A NEUROTRANSMITTER/
NEUROMODULATOR IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.
-!- TISSUE SPECIFICITY: HYPOTHALAMUS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M63582; AAA36480.1; -;
EMBL; M63581; AAA36480.1; JOINED.
PIR; A34550; A34550.
MIM; 275120; -;
Amidation; Hormone; Repeat; Hypothalamus; Signal;
Cleavage on pair of basic residues.
SIGNAL 1 724 POTENTIAL.
CHAIN 25 242 PROTHYROLIBERIN.
PEPTIDE 84 86 THYROLIBERIN.
PEPTIDE 114 116 THYROLIBERIN.
PEPTIDE 135 137 THYROLIBERIN.
PEPTIDE 152 154 THYROLIBERIN.
PEPTIDE 186 188 THYROLIBERIN.
PEPTIDE 227 229 THYROLIBERIN.
MOD_RES 84 84 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 86 86 AMIDATION (G-87 PROVIDE AMIDE GROUP).
MOD_RES 114 114 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 116 116 AMIDATION (G-117 PROVIDE AMIDE GROUP).
MOD_RES 135 135 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 137 137 AMIDATION (G-138 PROVIDE AMIDE GROUP).
MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 154 154 AMIDATION (G-155 PROVIDE AMIDE GROUP).
MOD_RES 186 186 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 188 188 AMIDATION (G-189 PROVIDE AMIDE GROUP).
MOD_RES 227 227 PYRROLIDONE CARBOXYLIC ACID.
MOD_RES 229 229 AMIDATION (G-230 PROVIDE AMIDE GROUP).
SEQUENCE 242 AA; 27404 MW; 80CF9D915B32F29F CRC64;

Query Match 0.6%; Score 8; DB 1; Length 242;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 169 EEEEEER 176
|||||||

RESULT 11
YNF8_YEAST STANDARD; PRT; 316 AA.
AC P53947;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 35.0 KDA PROTEIN IN ARP5-OMP2 INTERGENIC REGION.
GN YNL058C OR N2433 OR YNL2433C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1676;
RX MEDLINE=96021608; PubMed=8533472;
RA Bergez P., Doignon F., Crouzet M.;

"The sequence of a 44 420 bp fragment located on the left arm of
chromosome XIV from Saccharomyces cerevisiae.";
Yeast 11:967-974(1995).
[2]
ERRATUM.
RX MEDLINE=97060022; PubMed=8904343;
RA Bergez P., Doignon F., Crouzet M.;
RL Yeast 12:297-297(1996).
CC -!- SIMILARITY: TO YEAST YIL117C.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U12141; AAA99653.1; -;
DR EMBL; 271334; CAA95932.1; -;
DR SGD; S0005003; YNL058C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT DOMAIN 22 25 POLY-THR.
FT DOMAIN 62 65 POLY-SER.
FT DOMAIN 235 243 POLY-GLU.
SQ SEQUENCE 316 AA; 35046 MW; ADFAD73924279550 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 237 EEEEEER 244
|||||||

RESULT 12
RLA0_ARCFU STANDARD; PRT; 339 AA.
ID RLA0_ARCFU
AC Q28781;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ACIDIC RIBOSOMAL PROTEIN P0 HOMOLOG (L10E).
GN RPLP0 OR AFI491.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
R Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
OF E.COLI PROTEIN L10.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.

RC TISSUE-Eosinophil;
RA MEDLINE-91160746; PubMed=1705901;
RX Aoki I., Shindoh Y., Nishida T., Nakai S., Hong Y.-M., Mio M.,
RZ Saito T., Tasaka K.;
RT "Sequencing and cloning of the cDNA of guinea pig eosinophil major
RI basic protein.";
RN FEBS Lett. 279:330-334 (1991).
RZ [2]
RZ PARTIAL SEQUENCE.
RX MEDLINE-94092714; PubMed=8268206;
RA Hashimoto Y., Nagaoka I., Yamashita T.;
RZ "Purification of the antibacterial fragments of guinea-pig major
RI basic protein.";
RN Biochim. Biophys. Acta 1203:236-242 (1993).
RZ -1- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC
CC REACTIONS FROM INFLAMMATIONS, SINCE MBP IS CAPABLE OF RELEASING
CC HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF
CC BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.
CC -1- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC
CC GRANULE (CRYSTALLOID CORE).
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90251; BAA14291.1; -;
DR PIR; S13625; S13625.
DR InterPro; IPR001304; -;
DR InterPro; IPR002352; -;
DR Pfam; PF00059; Lectin_G; 1.
DR PRINTS; PR00770; EMABORBASCP.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
KW Lectin; Multigene family.
FT SIGNAL 1 15 POTENTIAL.
FT PROPEP 16 114 ACIDIC.
FT CHAIN 115 233 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1.
FT DOMAIN 132 233 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 134 231 BY SIMILARITY.
FT DISULFID 208 223 BY SIMILARITY.
SQ SEQUENCE 233 AA; 26268 MW; C8D5E96D927C56C8 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
DB 73 PEEEEEE 80

RESULT 9
SQML_ONCKE STANDARD; PRT; 233 AA.
AC P24405;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE SOMATOTROPIN PRECURSOR (SL).
OS Oncomorphus keta (Chum salmon).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
CC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8018;
RN [1]
RP SEQUENCE FROM N.A.
RA

RX MEDLINE-92017883; PubMed=1922096;
RA Takayama Y., Rand-Weaver M., Kawauchi H., Ono M.;
RT "Gene structure of chum salmon somatolactin, a presumed pituitary
RI hormone of the growth hormone/prolactin family.";
RN Mol. Endocrinol. 5:778-786 (1991).
RZ [2]
RZ SEQUENCE FROM N.A.
RX MEDLINE-92038879; PubMed=1936917;
RA Takayama Y., Ono M., Rand-Weaver M., Kawauchi H.;
RT "Greater conservation of somatolactin, a presumed pituitary hormone
RI of the growth hormone/prolactin family, than of growth hormone in
RN teleost fish.";
RZ Gen. Comp. Endocrinol. 83:366-374 (1991).
RZ -1- FUNCTION: MAY BE ASSOCIATED WITH ION REGULATION AND REPRODUCTION.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PITUITARY GLAND.
CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10640; BAA01487.1; -;
DR EMBL; D10638; BAA01485.1; -;
DR EMBL; D10636; BAA01485.1; JOINED.
DR EMBL; D10637; BAA01485.1; JOINED.
DR PIR; A23729; A23729.
DR PIR; B49762; B49762.
DR HSSP; P01241; IBP3.
DR InterPro; IPR001400; -;
DR Pfam; PF00103; hormone; 1.
DR PRINTS; PR00836; SOMATOTROPIN.
DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
KW Hormone; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 233 SOMATOTROPIN.
FT DISULFID 29 39 BY SIMILARITY.
FT DISULFID 89 205 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
SQ SEQUENCE 233 AA; 26882 MW; 178F385705F0A461 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 979 VSEKLSSL 986
DB 150 VSEKLSSL 157

RESULT 10
THYL_HUMAN STANDARD; PRT; 242 AA.
ID AC P20396;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE THYROLIBERIN PRECURSOR (THYROTROPIN RELEASING HORMONE) (TRH).
GN TRH.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91125361; PubMed=2126343;
RA Yamada M., Radovick S., Wondisford F.E., Nakayama Y., Weintraub B.D.,

KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
FT Oxidoreductase; Transit peptide.
FT TRANSIT 1 13
FT CHAIN 14 91
FT UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT 11 KDA PROTEIN.
FT DOMAIN 17 25
FT POLY-GLU.
SQ SEQUENCE 91 AA; 10755 MW; 499B44E100138F4C CRC64;

Query Match 0.6%; Score 8; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
D 16 PEEEEEE 23

RESULT 6
VE4_HPV04
ID VE4_HPV04 STANDARD; PRT; 181 AA.
AC 007852;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE PROBABLE E4 PROTEIN.
OS Human papillomavirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
OX NCBI_TaxID=10617;

SEQUENCE FROM N.A.
MEDLINE=93276568; PubMed=8389082;
Egawa K., Dellius H., Matsukura T., Kawashima M., de Villiers E.M.;
"Two novel types of human papillomavirus, HPV 63 and HPV 65:
comparisons of their clinical and histological features and DNA
sequences to other HPV types";
Virology 194:789-799(1993).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X70827; CAA50161.1; -
Early protein.
KW DOMAIN 133 142
SQ SEQUENCE 181 AA; 20849 MW; 950429799E00D5B CRC64;

Query Match 0.6%; Score 8; DB 1; Length 181;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
D 136 EEEEEER 143

RESULT 7
HMG2_CHICK
ID HMG2_CHICK STANDARD; PRT; 206 AA.
AC F26584;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN HMG2 (HMG-2).
GN HMG2.

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92241676; PubMed=1572546;
RA Davis D.L., Burch J.B.E.;
RT "Isolation of a chicken HMG2 cDNA clone and evidence for an HMG2-
specific 3'-untranslated region.";
RL Gene 113:251-256(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92290291; PubMed=1601311;
RA Sparrow D.B., Wells J.R.E.;
RT "Sequence of a cDNA encoding chicken high-mobility-group protein-2";
RL Gene 114:289-290(1992).
CC -!- FUNCTION: BINDS PREFERENTIALLY SINGLE-STRANDED DNA AND UNWINDS
DOUBLE STRANDED DNA.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HMG1/HMG2 PROTEIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 HMG BOXES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: M82335; AAA48818.1; -
DR EMBL: M80574; AAA48819.1; -
DR PIR: JC1114; JC1114.
DR PIR: JC1129; JC1129.
DR HSP; P07155; IHME.
DR InterPro; IPR000135; -
DR InterPro; IPR000910; -
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBLTY12.
DR PROSITE; PS00353; HMG1_2; 1.
KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
FT INIT_MET 0 0
FT DNA_BIND 8 78 HMG BOX 1.
FT DNA_BIND 94 162 HMG BOX 2.
FT DOMAIN 185 206 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 23 24 RE -> PR (IN REF. 2).
FT CONFLICT 136 136 L -> S (IN REF. 2).
SQ SEQUENCE 206 AA; 23696 MW; 2B058869208E27ID CRC64;

Query Match 0.6%; Score 8; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
D 186 PEEEEEE 193

RESULT 8
EMBL_CAVPO
ID EMBL_CAVPO STANDARD; PRT; 233 AA.
AC P22032;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1 PRECURSOR (MBP-1).
GN MBP1.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 115-162.

FT DOMAIN 841 1938 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 656 678 ACTIN-BINDING.
FT DOMAIN 758 772 ACTIN-BINDING.
FT DOMAIN 789 806 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 815 832 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 128 128 METHYLATION (SH-1) (POTENTIAL).
FT MOD_RES 696 696 METHYLATION (SH-2) (POTENTIAL).
FT MOD_RES 706 706 ALKYLATION (SH-1) (POTENTIAL).
FT CONFLICT 13 13 R -> AP (IN REF. 3).
FT CONFLICT 46 46 V -> A (IN REF. 3).
FT CONFLICT 51 52 VS -> AP (IN REF. 3).
FT CONFLICT 87 87 E -> Q (IN REF. 3).
FT CONFLICT 109 109 MISSING (IN REF. 3).
FT CONFLICT 1566 1566 F -> FF (IN REF. 4).
FT CONFLICT 1575 1575 R -> S (IN REF. 4).
FT CONFLICT 1721 1721 N -> T (IN REF. 4).
FT CONFLICT 1852 1852 T -> N (IN REF. 4).
FT CONFLICT 1870 1870 D -> N (IN REF. 4).
FT CONFLICT 1934 1934 M -> I (IN REF. 4 AND 5).
SQ SEQUENCE 1938 AA; 223507 MW; D7BD33FC2B19E3C2 CRC64;

Query Match 0.6%; Score 9; DB 1; Length 1938;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKOLEV 565
DQ 1528 EKIRKOLEV 1536
IIIIIIIIII

RESULT 4
GBG_LOLFO STANDARD; PRT; 87 AA.
AC 001821.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN GAMMA SUBUNIT.
OS Lolligo forbesi (Northern European squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
OC Myopsida; Loliginidae; Lolligo.
OX NCBI_TaxID=6618;
RN [1]
RE SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Retina;
RX MEDLINE=93050202; PubMed=1426257;
RA Lott J.S., Ryba N.J.P., Pottinger J.D.D., Keen J.N., Carne A.,
RA Findlay J.B.C.;
RT "The gamma-subunit of the principal G-protein from squid (Lolligo forbesi) photoreceptors contains a novel N-terminal sequence.";
RL FEBS Lett. 312:241-244(1992).
RS [2]
RH PRELIMINARY PARTIAL SEQUENCE OF 28-72.
RA MEDLINE=92028830; PubMed=1930153;
RA Pottinger J.D.D., Ryba N.J.P., Keen J.N., Findlay J.B.C.;
RT "The identification and purification of the heterotrimeric GTP-binding protein from squid (Lolligo forbesi) photoreceptors.";
RL Biochem. J. 279:323-326(1991).
RS [1]
RH INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE SIGNALING SYSTEMS. THIS MAJOR G-PROTEIN OF THE SQUID PHOTORECEPTOR IS INVOLVED IN VISUAL TRANSDUCTION. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-EFFECTOR INTERACTION.
CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -1- SIMILARITY: BELONGS TO THE G PROTEIN GAMMA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; M36647; AAA36317.1; -.
CC EMBL; Y00764; CAA68733.1; -.
CC PIR; S00219; S00219.

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; Z15112; CAA78816.1; -.
DR PIR; S25626; S25626.
DR PIR; S17596; S17596.
DR PIR; S27275; S27275.
DR InterPro; IPR001770;
DR PROSITE; PS00058; G-PROTEIN GAMMA; FALSE NEG.
KW Transducer; Prenylation; Lipoprotein; Vision.
FT DOMAIN 7 11 POLY-LYS.
FT DOMAIN 12 18 POLY-GLU.
FT MOD_RES 1 1 BLOCKED.
FT LIPID 84 84 GERANYL-GERANYL (BY SIMILARITY).
FT PROPEP 85 87 REMOVED IN MATURE FORM (BY SIMILARITY).
SQ SEQUENCE 87 AA; 10278 MW; BF9E07DEB780B104 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 87;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DQ 12 EEEEEER 19
IIIIIIIIII

RESULT 5
UCRH_HUMAN STANDARD; PRT; 91 AA.
AC P07919;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KDA PROTEIN PRECURSOR (EC 1.10.2.2) (MITOCHONDRIAL HINGE PROTEIN) (CYTOCHROME C1, NONHEME 11 KDA PROTEIN) (COMPLEX III SUBUNIT VIII).
GN UQCRH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RE SEQUENCE FROM N.A.
RX MEDLINE=88083627; PubMed=2826252;
RA Ohta S., Goto K., Arai H., Kagawa Y.;
RT "An extremely acidic amino-terminal presequence of the precursor for the human mitochondrial hinge protein.";
RL FEBS Lett. 226:171-175(1987).
RS [1]
RH FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY MEDIANTE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C -> Q + 2 FERROCYTOCHROME C.
CC -1- SUBUNIT: BCI COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND 6 LOW-MOLECULAR WEIGHT PROTEINS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE UQCRH/QCR6 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; M36647; AAA36317.1; -.
CC EMBL; Y00764; CAA68733.1; -.
CC PIR; S00219; S00219.

EMBL; M76599; AAA37160.1; -
 EMBL; M76600; AAA37161.1; -
 EMBL; M76601; AAA37162.1; -
 EMBL; M62404; AAA37424.1; -
 HSSP; P08799; LMND.
 SWISS-2DPAGE; Q02566; MOUSE.
 MGD; MGI:97255; Myhca.
 InterPro; IPR000048; -
 InterPro; IPR001609; -
 InterPro; IPR002928; -
 Pfam; PF00612; IQ; 1.
 Pfam; PF01576; Myosin_tail; 1.
 Pfam; PF00063; myosin_head; 1.
 PRINTS; PR00193; MYOSINHEAVY.
 PROSITE; PS50096; IQ; 1.
 Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
 Multigene family; Calmodulin-binding; Polymorphism.
 DOMAIN 1 841 GLOBULAR HEAD (S1).
 DOMAIN 842 1938 RODLIKE TAIL (S2 AND LMW DOMAINS).
 NP_BIND 842 1938 COILED COIL (POTENTIAL).
 DOMAIN 178 185 ATP.
 DOMAIN 657 679 ACTIN-BINDING.
 DOMAIN 759 773 ACTIN-BINDING.
 DOMAIN 790 807 CALMODULIN-BINDING (BY SIMILARITY).
 DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).
 MOD_RES 129 129 METHYLATION (TR-) (POTENTIAL).
 MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
 MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
 VARIANT 194 194 Y -> D.
 VARIANT 545 545 S -> A.
 VARIANT 838 838 I -> S.
 SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 0.6%; Score 9; DB 1; Length 1938;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

557 EKIRKQLEV 565
 |||||
 1529 EKIRKQLEV 1537

RESULT 3
 MESA_RAT STANDARD; PRT; 1938 AA.
 ID MYSA_RAT
 AC P02563; Q63351;
 DT 21-JUL-1986 (Rel. 01, Created)
 DE 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM.
 GN MYH6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 SEQUENCE FROM N.A.
 TISSUE=Heart;
 MEDLINE=90016822; PubMed=2798111;
 Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 "Complete nucleotide sequence of full length cDNA for rat alpha
 cardiac myosin heavy chain."
 Nucleic Acids Res. 17:7527-7528(1989).
 [2]
 DISCUSSION OF SEQUENCE.
 MEDLINE=90133919; PubMed=2614840;
 McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
 Comparisons suggest a molecular basis for functional differences."
 J. Mol. Biol. 210:665-671(1989).
 [3]

RP SEQUENCE OF 1-167 FROM N.A.
 RX MEDLINE=84194059; PubMed=6585819;
 RA Mahdavi V., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac alpha- and beta-myosin heavy chain genes are organized in
 tandem.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:2626-2630(1984).
 RN [4]
 RP SEQUENCE OF 1512-1938 FROM N.A.
 RX MEDLINE=82220036; PubMed=7045682;
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
 RT "Molecular characterization of two myosin heavy chain genes expressed
 in the adult heart."
 RL Nature 297:659-664(1982).
 RN [5]
 RP SEQUENCE OF 1872-1938 FROM N.A.
 RC STRAIN=Wistar; TISSUE=Heart;
 RX MEDLINE=85179510; PubMed=6241892;
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac myosin heavy chain isozymic transitions during development
 and under pathological conditions are regulated at the level of mRNA
 availability.";
 RL Eur. Heart J. 5:181-191(1984).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- MISCELLANEOUS: THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN
 THE RAT, TWO OF WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY
 CHAINS.
 CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
 WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
 CONSERVED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X15938; CAA34064.1; -
 DR EMBL; K01464; AAA41648.1; -
 DR EMBL; J00751; AAA41653.1; -
 DR EMBL; M32697; AAA41658.1; -
 DR PIR; S06005; S06005.
 DR PIR; A02988; A02988.
 DR HSSP; P08799; LMND.
 DR InterPro; IPR000048; -
 DR InterPro; IPR001609; -
 DR InterPro; IPR002928; -
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PROSITE; PS50096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
 KW Multigene family; Calmodulin-binding.
 FT DOMAIN 1 840 GLOBULAR HEAD (S1).
 FT DOMAIN 841 1938 RODLIKE TAIL (S2 AND LMW DOMAINS).
 FT FT

983 0.4 312 1 DBX-ANAPL O57314 anas platyr
984 0.4 312 1 YR81-CAEEL Q09566 caenorhabdi
985 0.4 313 1 NUI1-BRALA Q21000 branchiosto
986 0.4 313 1 YDJB-SCHPO P87059 schizosach
987 0.4 314 1 ACCO-MALDO Q00985 malus domes
988 0.4 314 1 LEUO-ECOLI P10151 escherichia
989 0.4 316 1 DHPS-STRPN P05382 streptococc
990 0.4 316 1 YVDE-BACSU Q06987 bacillus su
991 0.4 317 1 AAC4-DICDI P14198 dictyosteli
992 0.4 317 1 CC28-CANAL P43063 candida alb
993 0.4 318 1 CYPR-YEAST P25334 saccharomyc
994 0.4 318 1 Y051-NPVP P10309 orgyia pseu
995 0.4 318 1 Y211-AQUAE O68405 aquifex aeo
996 0.4 318 1 Y220-AQUAE O68411 aquifex aeo
997 0.4 319 1 VMP-CERV P05396 carnation e
998 0.4 321 1 IPNS-STRCT Q53932 streptomyce
999 0.4 321 1 Y4JO-RHISN P55515 rhizobium s
1000 0.4 322 1 SUCA-CAEEL P53596 caenorhabdi

ALIGNMENTS

RESULT 1
TRFL_HUMAN STANDARD; PRT: 439 AA.
AC P54274;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TELOMERIC REPEAT BINDING FACTOR 1.
GN TRFL OR TRFL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RE SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=96099400; PubMed=7502076;
RA Chong L., van Steensel B., Broccoli D., Erdjument-Bromage H.,
RA Hanish J., Tempst P., de Lange T.;
RA "A human telomeric protein.";
RL Science 270:1663-1667(1995).
RN [2]
RE REVISION TO 14.
RA de Lange T.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RE STRUCTURE BY NMR OF 378-430.
RA MEDLINE=98416700; PubMed=9739097;
RA Nishikawa T., Nagadoi A., Yoshimura S., Almoto S., Nishimura Y.;
RA "Solution structure of the DNA-binding domain of human telomeric
RA protein, hTRF1.";
RL Structure 6:1057-1065(1998).
CC -!- FUNCTION: BINDS THE TELOMERIC DOUBLE-STRANDED TTAGGG REPEAT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZES WITH TELOMERIC DNA IN
CC METAPHASE CELLS AND IS LOCATED AT CHROMOSOME ENDS DURING
CC METAPHASE.
CC -!- SIMILARITY: CONTAINS A MYB-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40705; -; NOT_ANNOTATED_CDS.
CC PDB; 1BA5; 27-APR-99.
CC MIM; 600951; -;
CC PROSITE; PS50090; MYB_3; 1.
CC DNA-binding; Telomere; Nuclear protein; 3D-structure.

FT DOMAIN 55 62 POLY-GLU.
FT DOMAIN 349 352 POLY-LYS.
FT DNA_BIND 375 428 MYB.
SQ SEQUENCE 439 AA: 50344 MW: 44FA3E7FF3E9CCAE CRC64;

Query Match 0.6%; Score 9; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331
IIIIIIIIII
Db 53 APEEEEEEE 61

RESULT 2
MYSA_MOUSE STANDARD; PRT: 1938 AA.
ID MYSA_MOUSE
AC Q02566; Q64258; Q64738;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM.
GN MYH6 OR MYHCA
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RA STRAIN=BALE/C. A/J, C57BL/6J, AND DBA/2J;
RX MEDLINE=92250040; PubMed=1577481;
RA Qulinn-Laquer B.K., Kennedy J.E., Wei S.J., Beisel K.W.;
RA "Characterization of the allelic differences in the mouse cardiac
RA alpha-myosin heavy chain coding sequence.";
RL Genomics 13:176-188(1992).
RN [2]
RE SEQUENCE OF 1-67 FROM N.A.
RC STRAIN=AKR;
RX MEDLINE=91225025; PubMed=2026617;
RA Gulick J., Subramaniam A., Neumann J., Robbins J.;
RA "Isolation and characterization of the mouse cardiac myosin heavy
RA chain genes.";
RL J. Biol. Chem. 266:9180-9185(1991).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M76598; AAA37159.1; -;

837	6	0.4	264	1	PPA_ZYMO	1	P14924	zymomonas m	910	6	0.4	291	1	CD20_MOUSE	P19437	mus musculus
838	6	0.4	265	1	TRPC_ZYMP	1	Q9xhm3	zymomonas m	911	6	0.4	291	1	GLD2_XENLA	P53782	xenopus lae
839	6	0.4	265	1	YCFH_ECOLI	1	P37346	escherichia	912	6	0.4	291	1	FLAV_MATIN	O04395	matthiola i
840	6	0.4	266	1	DSHB_YEAST	1	P21801	saccharomyc	913	6	0.4	292	1	AMPR_PROST	O69772	providencia
841	6	0.4	266	1	NIFM_KLEPN	1	P08534	klebsiella	914	6	0.4	292	1	BIP2_TOBAC	Q03682	nicotiana t
842	6	0.4	266	1	YD37_SCHPO	1	Q10271	schizosacch	915	6	0.4	292	1	CC22_ORYSA	P29619	oryza sativ
843	6	0.4	267	1	143H_ARATH	1	P46077	arabidopsis	916	6	0.4	292	1	P34117 dictyosteli	P34117	dictyosteli
844	6	0.4	267	1	CGIC_DROME	1	P25008	drosophila	917	6	0.4	293	1	TYRC_ZYMO	Q04983	zymomonas m
845	6	0.4	267	1	PENK_HUMAN	1	P01210	homo sapien	918	6	0.4	294	1	ARR1_YEAST	P06596	saccharomyc
846	6	0.4	267	1	PKI_NPVH2	1	P41719	heliobis z	919	6	0.4	294	1	CC21_ORYSA	P29618	oryza sativ
847	6	0.4	267	1	SUHE_HAEIN	1	P44333	haemophilus	920	6	0.4	294	1	CC22_MEDSA	Q05006	medicago sa
848	6	0.4	268	1	CCHL_HUMAN	1	P53701	homo sapien	921	6	0.4	294	1	CC2A_ANTMA	Q38772	antirrhinum
849	6	0.4	268	1	ILIA_BOVIN	1	P08831	bos taurus	922	6	0.4	294	1	CC2A_ARATH	P24100	arabidopsis
850	6	0.4	268	1	ILIA_CAPHI	1	P79161	capra hircu	923	6	0.4	294	1	CC2_MAIZE	P23111	zea mays (m
851	6	0.4	268	1	ILIA_SHEEP	1	Q28579	ovis aries	924	6	0.4	294	1	CC2_VIGUN	P52389	vigna unqui
852	6	0.4	268	1	INHA_MYCAV	1	O07400	mycobacteri	925	6	0.4	294	1	SMN_HUMAN	Q16637	homo sapien
853	6	0.4	268	1	ISPE_HELPY	1	O25984	helicobacte	926	6	0.4	295	1	CPPW_SALTY	Q56027	salmonella
854	6	0.4	268	1	RAD_RAT	1	P55043	rattus norv	927	6	0.4	295	1	MAUJ_METEX	Q49127	methylobact
855	6	0.4	268	1	TRPA_ECOLI	1	P00928	escherichia	928	6	0.4	296	1	EGPM_ECOLI	P77541	escherichia
856	6	0.4	268	1	TRPA_SALTY	1	P00929	salmonella	929	6	0.4	296	1	EGR3_MOUSE	P43300	mus musculu
857	6	0.4	268	1	VANY_ENTFA	1	Q47746	enterococcu	930	6	0.4	296	1	THTM_RAT	P97532	rattus norv
858	6	0.4	269	1	RAD_HUMAN	1	P55042	homo sapien	931	6	0.4	296	1	YBNB_ECOLI	P77395	escherichia
859	6	0.4	269	1	TRPA_KLEAE	1	P00930	klebsiella	932	6	0.4	297	1	CD20_HUMAN	P11836	homo sapien
860	6	0.4	269	1	VC03_SPVKA	1	P32229	swinepox vi	933	6	0.4	297	1	DSR3_HUMAN	O14972	homo sapien
861	6	0.4	269	1	VG01_HSVSA	1	Q01036	herpesvirus	934	6	0.4	297	1	DSR3_MOUSE	O35075	mus musculu
862	6	0.4	270	1	CDSA_BRUAB	1	Q59173	b phosphati	935	6	0.4	297	1	MYOD_COTJA	P21572	coturnix co
863	6	0.4	270	1	ILIA_HORSE	1	Q28385	equus cabal	936	6	0.4	297	1	RRPP_RABVA	P15198	rabies viru
864	6	0.4	270	1	ILIA_MOUSE	1	P01582	mus musculu	937	6	0.4	297	1	RRPP_RABVC	P22363	rabies viru
865	6	0.4	270	1	PROC_CORGL	1	P46340	corynebacte	938	6	0.4	297	1	RRPP_RABVE	P22559	rabies viru
866	6	0.4	271	1	TRPA_MYCIT	1	O68906	mycobacteri	939	6	0.4	297	1	RRPP_RABVP	P06747	rabies viru
867	6	0.4	271	1	YAHL_ECOLI	1	P77393	escherichia	940	6	0.4	297	1	RRPP_RABVS	P16286	rabies viru
868	6	0.4	271	1	YG5W_YEAST	1	P53335	saccharomyc	941	6	0.4	297	1	VGLG_HRSVA	P27023	human respi
869	6	0.4	271	1	DPQI_KLEPN	1	P26509	klebsiella	942	6	0.4	297	1	VGLG_HRSV6	P27025	human respi
870	6	0.4	272	1	HKX1_STRVN	1	P17542	streptomyce	943	6	0.4	297	1	VGLG_HRSV7	P27026	human respi
871	6	0.4	272	1	VALI_MSVK	1	P03368	maize strea	944	6	0.4	298	1	CC28_YEAST	P00546	saccharomyc
872	6	0.4	272	1	VALI_MSVN	1	P14980	maize strea	945	6	0.4	298	1	MMSB_PSEAE	P28811	pseudomonas
873	6	0.4	272	1	VALI_MSVS	1	P14590	maize strea	946	6	0.4	298	1	MYOD_CHICK	P16075	gallus gall
874	6	0.4	273	1	OSAI_BORBU	1	P14013	borrelia bu	947	6	0.4	298	1	VGLG_HRSVA	P27024	human respi
875	6	0.4	273	1	T2C2_HERAU	1	P25959	herpetosiph	948	6	0.4	298	1	VGLG_HRSVA	P03423	human respi
876	6	0.4	273	1	TCJA_CAEBR	1	P35072	caenorhabdi	949	6	0.4	298	1	VGLG_HRSVL	P20895	human respi
877	6	0.4	273	1	TCJA_CAEBR	1	Q04202	caenorhabdi	950	6	0.4	299	1	YF02_MYCTU	P17183	mycobacteri
878	6	0.4	274	1	PSBS_SPIOL	1	Q02060	spinacia ol	951	6	0.4	299	1	YYAK_BACSU	P37513	bacillus su
879	6	0.4	275	1	AMOQ_NITEU	1	Q44507	nitrosomona	952	6	0.4	300	1	CPPM_BACHD	Q929t7	bacillus ha
880	6	0.4	275	1	CDID_NITFL	1	P23043	syllilagug	953	6	0.4	300	1	NIFU_ANAAZ	Q43885	anabaena az
881	6	0.4	275	1	LEC_PEA	1	P02867	pisum sativ	954	6	0.4	300	1	SPRC_XENLA	P36378	xenopus lae
882	6	0.4	275	1	QJQG_BACSU	1	P54544	bacillus su	955	6	0.4	300	1	T2D3_HABIN	P43870	haemophilus
883	6	0.4	277	1	DLQH_RHET	1	O05940	rhizobium e	956	6	0.4	301	1	KIPN_BPFA4	P06855	bacterioph
884	6	0.4	279	1	HAXI_HUMAN	1	O00165	homo sapien	957	6	0.4	301	1	MCP_BPFA1	P26596	lactococcus
885	6	0.4	279	1	THEI_THEVU	1	P04072	thermoactin	958	6	0.4	301	1	OPSD_CAMSC	O16018	cambarellus
886	6	0.4	280	1	AROE_BACSU	1	P54374	bacillus su	959	6	0.4	301	1	SCRK_ZYMO	Q03417	zymomonas m
887	6	0.4	280	1	CC2B_ANTMA	1	Q38773	antirrhinum	960	6	0.4	301	1	TRT2_CHICK	P02642	gallus gall
888	6	0.4	280	1	DNAT_TETH	1	Q56237	thermus aqu	961	6	0.4	302	1	FABD_MYCTU	Q10501	mycobacteri
889	6	0.4	281	1	RP32_HAEIN	1	P44404	haemophilus	962	6	0.4	306	1	DHB3_RAT	Q34939	rattus norv
890	6	0.4	281	1	TRPA_METJA	1	Q60180	methanococc	963	6	0.4	306	1	MPR3_YEAST	P43588	saccharomyc
891	6	0.4	281	1	YTXC_BACSU	1	P06569	bacillus su	964	6	0.4	306	1	OPPB_ECOLI	P31132	escherichia
892	6	0.4	282	1	PANC_CAMJE	1	Q9plk2	campylobact	965	6	0.4	306	1	OPPB_SALTY	P08005	salmonella
893	6	0.4	283	1	FLAB_LEPIN	1	O51941	leptosira	966	6	0.4	306	1	YBFH_BACSU	Q31448	bacillus su
894	6	0.4	283	1	RP38_HUMAN	1	P78345	homo sapien	967	6	0.4	307	1	YDFB_BACSU	P96680	bacillus su
895	6	0.4	284	1	RP32_PROMI	1	P50509	proteus mir	968	6	0.4	307	1	SRB5_YEAST	P32585	saccharomyc
896	6	0.4	284	1	YURL_BACSU	1	Q32153	bacillus su	969	6	0.4	308	1	FRP_HUMAN	Q12841	homo sapien
897	6	0.4	285	1	RP32_ENTCL	1	P50508	enterobacte	970	6	0.4	308	1	KHSE_BRELA	P07128	brevibacter
898	6	0.4	285	1	YAT1_SCHPO	1	Q10148	schizosacch	971	6	0.4	308	1	KHSE_CORGL	P08210	corynebacte
899	6	0.4	285	1	YD98_YEAST	1	Q03835	saccharomyc	972	6	0.4	308	1	MENA_HABIN	P44739	haemophilus
900	6	0.4	287	1	FOLD_SALTI	1	Q60006	salmonella	973	6	0.4	308	1	YBIN_ECOLI	P75782	escherichia
901	6	0.4	287	1	MF25_XENLA	1	P16076	xenopus lae	974	6	0.4	310	1	EL3A_HORVU	P37422	hordeum vul
902	6	0.4	288	1	DSR2_HUMAN	1	O95456	homo sapien	975	6	0.4	310	1	GRB_BP22	P57022	bacterioph
903	6	0.4	288	1	ILVE_METJA	1	Q58414	methanococc	976	6	0.4	310	1	HEM2_BACHD	Q9kdk9	bacillus ha
904	6	0.4	288	1	LIPH_PSEAE	1	Q01725	pseudomonas	977	6	0.4	310	1	PURU_MYCTU	Q50453	mycobacteri
905	6	0.4	290	1	CAH8_HUMAN	1	P35219	homo sapien	978	6	0.4	310	1	VU84_HSV7J	P52534	human herpe
906	6	0.4	290	1	GSFC_HERSA	1	P45772	aeromonas s	979	6	0.4	310	1	Y4RE_RHLSN	P55638	rhizobium s
907	6	0.4	290	1	TRX2_MOUSE	1	O08550	mus musculu	980	6	0.4	310	1	YRRI_EBV	P03207	epstein-bar
908	6	0.4	291	1	ATPG_BUDSA	1	O51873	buchnera ap	981	6	0.4	311	1	ATH6_ARATH	P46668	arabidopsis
909	6	0.4	291	1	CC21_MEDSA	1	P24923	medicago sa	982	6	0.4	312	1	ATH5_ARATH	P46667	arabidopsis

691	1	196	0.4	6	028220	archaeoglob	764	6	0.4	233	1	RR2_CVACA	P35014	cyaniidum c
692	1	196	0.4	6	P38472	marchantia	765	6	0.4	233	1	YB63_METJA	Q58858	methanococc
693	1	197	0.4	6	P05422	phyllomedus	766	6	0.4	234	1	MAG8_HUMAN	P43361	homo sapien
694	1	198	0.4	6	P05421	phyllomedus	767	6	0.4	234	1	SAMP_MESAU	P07629	mesocricetu
695	1	199	0.4	6	O32481	legionella	768	6	0.4	234	1	SNG1_RAT	Q62876	rattus norv
696	1	199	0.4	6	O05386	bacillus ce	769	6	0.4	234	1	YJ72_YEAST	P39543	saccharomyc
697	1	199	0.4	6	O84971	rhodobacter	770	6	0.4	235	1	END1_ECOLI	P25736	escherichia
698	1	201	0.4	6	P00106	pseudomonas	771	6	0.4	235	1	FLIH_SALTY	P15934	salmonella
699	1	201	0.4	6	P27344	saccharomyc	772	6	0.4	235	1	GAMT_RAT	P10868	rattus norv
700	1	202	0.4	6	O04905	arabidopsis	773	6	0.4	235	1	MTN_BUCAP	O51931	buchnera ap
701	1	203	0.4	6	P15771	homo sapien	774	6	0.4	235	1	PYG_AGLNE	P34814	aglaothamni
702	1	203	0.4	6	P53288	saccharomyc	775	6	0.4	235	1	TFE2_ALCEU	P94136	alcaligenes
703	1	204	0.4	6	P45273	haemophilus	776	6	0.4	235	1	VINC_XENLA	Q04615	xenopus lae
704	1	204	0.4	6	Q05163	dicentrarch	777	6	0.4	236	1	CYSA_CHLVU	P56344	chlorella v
705	1	204	0.4	6	Q01283	lates calca	778	6	0.4	236	1	GAMT_HUMAN	P14353	homo sapien
706	1	204	0.4	6	P48248	morone saxa	779	6	0.4	236	1	GAMT_MOUSE	O35969	mus musculus
707	1	204	0.4	6	P09539	seriola qui	780	6	0.4	236	1	PLC1_BOVIN	P09611	bos taurus
708	1	204	0.4	6	P09113	thunnus thy	781	6	0.4	237	1	NDF3_HUMAN	O92886	homo sapien
709	1	204	0.4	6	Q16890	homo sapien	782	6	0.4	238	1	CHVI_RHISN	P50351	rhizobium s
710	1	204	0.4	6	O54818	mus musculus	783	6	0.4	239	1	DCUR_ECOLI	P39271	escherichia
711	1	204	0.4	6	Q06600	bos taurus	784	6	0.4	239	1	UPPS_HAEIN	P44938	haemophilus
712	1	205	0.4	6	P34555	caenorhabdi	785	6	0.4	240	1	CHVI_RHIME	P50350	rhizobium m
713	1	206	0.4	6	P05205	drosophila	786	6	0.4	240	1	HDFE_HUMAN	P51858	homo sapien
714	1	206	0.4	6	Q06630	manduca sex	787	6	0.4	240	1	PYRH_BACSU	O31749	baecillus su
715	1	206	0.4	6	Q58298	methanococc	788	6	0.4	241	1	AGL8_SINAL	Q41274	sinapius alb
716	1	206	0.4	6	P76264	escherichia	789	6	0.4	241	1	CHVI_AGRTO	Q07783	agrobacteri
717	1	208	0.4	6	P28074	homo sapien	790	6	0.4	241	1	PHNF_ECOLI	P16684	escherichia
718	1	208	0.4	6	Q12488	saccharomyc	791	6	0.4	242	1	AGL8_ARATH	Q38876	arabidopsis
719	1	209	0.4	6	Q04603	human immu	792	6	0.4	243	1	ARTJ_ECOLI	Q30860	escherichia
720	1	209	0.4	6	O51560	borrelia bu	793	6	0.4	243	1	TONB_ENTAE	P46383	enterobacte
721	1	210	0.4	6	P27682	rattus norv	794	6	0.4	243	1	TPX1_HUMAN	P16562	homo sapien
722	1	210	0.4	6	O91814	gallus gall	795	6	0.4	243	1	TRIC_XENLA	P50754	xenopus lae
723	1	210	0.4	6	P51560	vibrio angu	796	6	0.4	244	1	SANA_HAEIN	P45130	haemophilus
724	1	210	0.4	6	P75256	mycoplasma	797	6	0.4	244	1	YKL4_CAEEL	P42171	caenorhabdi
725	1	210	0.4	6	Q11071	caenorhabdi	798	6	0.4	245	1	CRS3_HORSE	O19010	equus cabal
726	1	211	0.4	6	P26487	azorhizobiu	799	6	0.4	245	1	NISL_LACLA	P42708	lactococcus
727	1	211	0.4	6	O08849	mus musculus	800	6	0.4	245	1	TM4B_HUMAN	Q94uk8	homo sapien
728	1	211	0.4	6	P33407	oryza sativ	801	6	0.4	245	1	YBPA_BURCE	P37335	burkholderi
729	1	211	0.4	6	P34452	caenorhabdi	802	6	0.4	246	1	CHID_LYCES	Q05537	lycopersico
730	1	211	0.4	6	Q04550	southampton	803	6	0.4	246	1	INTL_ECOLI	Q47036	escherichia
731	1	211	0.4	6	P45390	escherichia	804	6	0.4	246	1	INTM_ECOLI	P56979	escherichia
732	1	212	0.4	6	P12961	mus musculus	805	6	0.4	246	1	MYP0_HETFR	P20938	heterodontu
733	1	212	0.4	6	P09457	saccharomyc	806	6	0.4	246	1	YBEC_CLOHI	Q92nk0	clostridium
734	1	212	0.4	6	Q02130	lactococcus	807	6	0.4	247	1	GRPB_RAT	P08462	rattus norv
735	1	213	0.4	6	P29227	drosophila	808	6	0.4	247	1	VGS_BPPI	P22588	bacterioph
736	1	213	0.4	6	O94468	schizosacch	809	6	0.4	248	1	DJ_DRONE	O01352	drosophila
737	1	214	0.4	6	P14204	bacillus su	810	6	0.4	250	1	AMPM_MYCGA	O52353	mycoplasma
738	1	214	0.4	6	P92847	bothriopsis	811	6	0.4	252	1	AGSD_ECOLI	P46004	escherichia
739	1	214	0.4	6	P22378	simian immu	812	6	0.4	252	1	DHSB_SCHPO	P21911	schizosacch
740	1	215	0.4	6	Q00717	emericeella	813	6	0.4	252	1	DUT_HUMAN	P33316	homo sapien
741	1	215	0.4	6	P75849	escherichia	814	6	0.4	252	1	NUOE_MYCTU	P95177	mycobacteri
742	1	216	0.4	6	P77985	staphylococ	815	6	0.4	252	1	PRGK_SALTY	P41786	salmonella
743	1	217	0.4	6	O60046	neurospora	816	6	0.4	254	1	COBJ_PSEDE	P21640	pseudomonas
744	1	217	0.4	6	P28161	homo sapien	817	6	0.4	255	1	ISPE_CAMUE	Q9pn10	campylobact
745	1	217	0.4	6	P09629	homo sapien	818	6	0.4	255	1	SKGA_CAUCR	Q9rpe7	caulobacter
746	1	217	0.4	6	P09024	mus musculus	819	6	0.4	256	1	ADH_DROTS	P51550	drosophila
747	1	217	0.4	6	P54601	bacillus su	820	6	0.4	256	1	OBFI_MOUSE	Q64693	mus musculus
748	1	218	0.4	6	P91580	ciona intes	821	6	0.4	257	1	MSRA_BRANA	P54151	brassica na
749	1	219	0.4	6	P54536	bacillus su	822	6	0.4	257	1	PTC1_CYNPY	O42335	cynops pyrr
750	1	220	0.4	6	Q91771	xenopus lae	823	6	0.4	258	1	MSRA_ARATH	P54150	arabidopsis
751	1	220	0.4	6	P37502	bacillus su	824	6	0.4	258	1	UL34_HSV7J	P52466	human herpe
752	1	221	0.4	6	P18432	drosophila	825	6	0.4	258	1	Y875_SYNY3	P73555	synecocyst
753	1	221	0.4	6	P19972	pichia fari	826	6	0.4	258	1	YO33_BPHP1	P51737	bacterioph
754	1	222	0.4	6	P33999	escherichia	827	6	0.4	259	1	MSRA_LACSA	Q9sec2	lactuca sat
755	1	226	0.4	6	P78035	mycoplasma	828	6	0.4	260	1	CABV_MOUSE	P12658	mus musculus
756	1	226	0.4	6	P04978	bordetella	829	6	0.4	261	1	MURI_BORBU	O51127	borrelia bu
757	1	226	0.4	6	P42399	bacillus su	830	6	0.4	261	1	SPRE_HUMAN	P35270	homo sapien
758	1	228	0.4	6	P31068	escherichia	831	6	0.4	261	1	YH86_SYNY3	P73335	synecocyst
759	1	228	0.4	6	P50540	mus musculus	832	6	0.4	262	1	PH4H_PSEAE	P43334	pseudomonas
760	1	231	0.4	6	O45551	caenorhabdi	833	6	0.4	262	1	SPRE_RAT	P18297	rattus norv
761	1	231	0.4	6	P26717	homo sapien	834	6	0.4	262	1	TRT3_CHICK	P12620	gallus gall
762	1	233	0.4	6	Q58309	methanococc	835	6	0.4	262	1	Y877_MYCTU	Q10539	mycobacteri
763	1	233	0.4	6	P26715	homo sapien	836	6	0.4	263	1	KDGR_ECOLI	P76268	escherichia

545	125	1	H2B_CHICK	P02279 gallus gall	618	6	0.4	162	1	THY1_MOUSE	P01831 mus musculu
546	127	1	RL22_XENLA	P50886 xenopus lae	619	6	0.4	163	1	ILVH_HAEIN	P45260 haemophilus
547	127	1	VYV1_AZOVI	P40431 azotobacter	620	6	0.4	164	1	ATPX_CVAPA	P48085 cyanophora
548	128	1	CHEY_ECOLI	P06143 escherichia	621	6	0.4	164	1	VG59_BFMLS	Q05272 mycobacteri
549	128	1	CHEY_SALTY	P06657 salmonella	622	6	0.4	165	1	LE24_ARCFU	O28513 archaeoglob
550	128	1	YT24_CABEL	Q10933 caenorhabdi	623	6	0.4	165	1	MLCH_HUMAN	P20382 homo sapien
551	129	1	A2UR_ALCXX	P56547 alcaligenes	624	6	0.4	165	1	MLCH_RAT	P14200 rattus norv
552	129	1	A2UR_ALCSP	P00279 alcaligenes	625	6	0.4	166	1	BIOX_BACSH	P22821 bacillus sp
553	129	1	RK12_GUTH	O78414 guillardia	626	6	0.4	166	1	ING_CAPHI	P79154 capra hircu
554	129	1	YASD_LEPIN	P41395 leptospira	627	6	0.4	166	1	MLCH_MOUSE	P56942 mus musculu
555	130	1	ACR2_YEAST	Q06597 saccharomyc	628	6	0.4	167	1	RNBR_BOVIN	P39873 bos taurus
556	130	1	LV1G_HUMAN	P06316 homo sapien	629	6	0.4	168	1	GA15_MOUSE	P35639 mus musculu
557	130	1	TAT_HVZBE	P18098 human immun	630	6	0.4	168	1	IA03_WHEAT	P17314 triticum ae
558	130	1	TVA1_MOUSE	P01738 mus musculu	631	6	0.4	168	1	TCTP_HVBR	Q92599 hevea bras
559	131	1	LYGE_HUMAN	Q16553 homo sapien	632	6	0.4	168	1	TCTP_TOBAC	O9xhl7 nicotiana t
560	131	1	PSBU_SINY3	Q55332 synechocyst	633	6	0.4	168	1	YE92_HAEIN	P44217 haemophilus
561	132	1	FLIS_PSEAE	O33422 pseudomonas	634	6	0.4	169	1	CAFI_DICDI	P54670 dictyostell
562	132	1	Y433_METH	O26533 methanobact	635	6	0.4	171	1	IAAD_HORVU	P11643 hordeum vul
563	133	1	TATB_VIBCH	P57063 vibrio chol	636	6	0.4	172	1	RL11_TETTH	P24119 tetrahymena
564	133	1	VG4_SPV4	P11336 spiroplasma	637	6	0.4	173	1	FRI3_RANCA	P07797 rana catesb
565	134	1	H2AV_YEAST	Q12692 saccharomyc	638	6	0.4	174	1	PLU_DROME	P42570 drosophila
566	134	1	IL5_CAVPO	O08987 cavia porce	639	6	0.4	175	1	AXIS_ARATH	Q9xfm0 arabidopsis
567	135	1	RS6_ECOLI	P02358 escherichia	640	6	0.4	176	1	RL10_STRCO	P41103 streptomyce
568	137	1	MP12_LYMT	P25289 lymanaea sta	641	6	0.4	177	1	GRPE_THETH	Q56236 thermus aqu
569	137	1	RETA_HUMAN	P29373 homo sapien	642	6	0.4	177	1	YNIW_AZCH	P23177 azotobacter
570	137	1	RK16_ODOSI	P49553 odontella s	643	6	0.4	178	1	LACB_PIG	P04119 sus scrofa
571	138	1	MP15_YEAST	P31241 lymanaea sta	644	6	0.4	179	1	FLIN_AGRU	Q57259 agrobacteri
572	138	1	YJH5_YEAST	P47036 saccharomyc	645	6	0.4	179	1	GRPE_LACLC	P42369 lactococcus
573	139	1	RISB_METH	O27443 methanobact	646	6	0.4	179	1	SPI7_MONDO	O62771 monodelphis
574	140	1	DHSB_ARATH	P21915 arabidopsis	647	6	0.4	179	1	UBC2_CANAL	O74201 candida alb
575	140	1	FUCU_ECOLI	P11555 escherichia	648	6	0.4	180	1	GLUC_CAVPO	P05110 cavia porce
576	140	1	YP15_STAAU	P13977 staphylococ	649	6	0.4	180	1	YAF5_YEAST	P39718 saccharomyc
577	141	1	ALG2_TYRPU	O02380 tyrophagus	650	6	0.4	182	1	HSB2_HUMAN	Q16082 homo sapien
578	141	1	YORB_CAEEL	Q09338 caenorhabdi	651	6	0.4	182	1	HSB2_RAT	Q35878 rattus norv
579	141	1	ISOD_SULAC	P37820 sulfolobus	652	6	0.4	182	1	NODC_BRAEL	P53417 bradyrhizob
580	143	1	IF2B_METJA	Q57562 methanococc	653	6	0.4	182	1	Y391_HAEIN	P43992 haemophilus
581	144	1	YB67_PSEAE	O68560 pseudomonas	654	6	0.4	182	1	YGFA_ECOLI	P09160 escherichia
582	145	1	HBBI_TRICR	P10785 triturus cr	655	6	0.4	183	1	E321_ADELA	P37169 human adeno
583	146	1	OR_CANPA	Q02720 canis famil	656	6	0.4	183	1	VATE_METMA	Q60183 methanosarc
584	147	1	MCP_MEDSA	P42067 medicago sa	657	6	0.4	183	1	YGJV_ECOLI	P42603 escherichia
585	147	1	SPAM_SALTY	P40612 salmonella	658	6	0.4	184	1	BCM_HUMAN	Q02223 homo sapien
586	147	1	UCRH_YEAST	P00127 saccharomyc	659	6	0.4	184	1	KITH_THEMA	Q9WYN2 thermotoga
587	147	1	Y846_SINY3	Q55433 synechocyst	660	6	0.4	185	1	CBX1_HUMAN	P23197 homo sapien
588	147	1	YI80_SINY3	P74125 synechocyst	661	6	0.4	185	1	FLAV_AQUAE	Q67866 aquifex aeo
589	148	1	BENE_HUMAN	Q13021 homo sapien	662	6	0.4	185	1	PHNN_ECOLI	P16690 escherichia
590	148	1	CALM_PARTE	P07463 paramecium	663	6	0.4	185	1	RR5_ODOSI	P49493 odontella s
591	148	1	CC21_PEA	P19026 pisum sativ	664	6	0.4	185	1	Y754_METJA	Q58164 methanococc
592	148	1	PFDA_PYRHO	O58263 pyrococcus	665	6	0.4	187	1	ATPD_ODOSI	Q00821 odontella s
593	148	1	YCBG_HAEIN	P44161 haemophilus	666	6	0.4	187	1	E321_ADE1P	P35770 human adeno
594	149	1	NDK_SINY3	P74494 synechocyst	667	6	0.4	187	1	E321_ADE35	P15139 human adeno
595	149	1	YFEZ_ECOLI	P76538 escherichia	668	6	0.4	189	1	ICVA_MANSE	P00305 manduca sex
596	150	1	YP59_YEAST	Q02784 saccharomyc	669	6	0.4	189	1	KADL_METTH	O27699 methanobact
597	151	1	YH20_RHOCA	Q02999 rhodobacter	670	6	0.4	189	1	RASK_HUMAN	P01116 homo sapien
598	152	1	RECX_VIBCH	Q56647 vibrio chol	671	6	0.4	189	1	RASK_MOUSE	P32883 mus musculu
599	152	1	YF47_SINY3	P74662 synechocyst	672	6	0.4	189	1	RASK_MSVKI	P01117 kirsten mur
600	152	1	YFJX_ECOLI	P52139 escherichia	673	6	0.4	189	1	YADM_ECOLI	P37018 escherichia
601	153	1	Y264_METJA	Q57712 methanococc	674	6	0.4	191	1	CLUS_MESAU	P14683 mesocricetu
602	154	1	KMLS_MELGA	P56276 meleagris g	675	6	0.4	191	1	SNGI_MOUSE	Q55100 mus musculu
603	154	1	YKOL_CAEEL	P34289 caenorhabdi	676	6	0.4	191	1	TF2D_PYRAB	Q9V024 pyrococcus
604	155	1	PCP_HAEIN	P10325 haemophilus	677	6	0.4	191	1	TF2D_PYRFU	O58750 pyrococcus
605	155	1	YHCH_HAEIN	P44583 haemophilus	678	6	0.4	191	1	TF2D_PYRHO	O58737 pyrococcus
606	156	1	RM25_YEAST	P23369 saccharomyc	679	6	0.4	193	1	GNTK_SCHPO	Q10242 schizosacch
607	156	1	UBCA_YEAST	P52492 saccharomyc	680	6	0.4	193	1	NUIC_SINY3	P26525 synechocyst
608	157	1	UBC3_SCHPO	P40984 schizosacch	681	6	0.4	193	1	Y377_MYCPN	P75223 mycoplasma
609	158	1	RHM2_HUMAN	P25791 homo sapien	682	6	0.4	193	1	Y455_METJA	Q57897 methanococc
610	158	1	RHM2_MOUSE	P25801 mus musculu	683	6	0.4	193	1	C24A_HUMAN	P13498 homo sapien
611	158	1	RL15_AERPE	Q9yf98 aeteryum p	684	6	0.4	194	1	PTH_ECOLI	P23932 escherichia
612	158	1	YIM2_BPH1	P10426 bacterioph	685	6	0.4	194	1	PTH_SALTI	Q60001 salmonella
613	159	1	PTXA_MYCPN	P75292 mycoplasma	686	6	0.4	194	1	Y041_THEAC	P57674 thermoplas
614	159	1	YHAG_YEAST	P38751 saccharomyc	687	6	0.4	194	1	YD16_HAEIN	P44159 haemophilus
615	161	1	THY1_RAT	P01830 rattus norv	688	6	0.4	196	1	NODA_RHIS3	P72329 rhizobium s
616	161	1	UREE_PROMI	P17090 proteus mir	689	6	0.4	196	1	NUDH_HAEIN	Q57045 haemophilus
617	161	1	Y0UD_CAEEL	P30631 caenorhabdi	690	6	0.4	196	1	RGS1_HUMAN	Q08116 homo sapien

399	7	0.5	1137	1	MSB1_YEAST	P21339	saccharomyc	472	6	0.4	76	1	ULD1_HCMVA	P16773	human cytom
400	7	0.5	1141	1	CN3A_HUMAN	P14432	homo sapien	473	6	0.4	77	1	HEMZ_YERPS	Q05338	yersinia ps
401	7	0.5	1141	1	CN3A_RAT	Q62865	rattus norv	474	6	0.4	79	1	Y437_METJA	Q57879	methanococc
402	7	0.5	1152	1	YK00_CAEEL	P34305	caenorhabdi	475	6	0.4	82	1	YFTD_METFE	P65059	methanother
403	7	0.5	1157	1	YXNA_THESA	P35917	thermoanaer	476	6	0.4	83	1	CALD_BOVIN	Q27976	bos taurus
404	7	0.5	1161	1	NRDC_RAT	P47245	rattus norv	477	6	0.4	83	1	CLSS_HAEMA	Q25088	haemophilus ma
405	7	0.5	1181	1	P1B2_HUMAN	Q00722	homo sapien	478	6	0.4	84	1	Y073_NPVOP	O10326	orgyia pseu
406	7	0.5	1214	1	BRF3_HUMAN	Q9ulda	homo sapien	479	6	0.4	87	1	EFTS_ANTSP	Q02855	antithamnio
407	7	0.5	1215	1	S1P1_MOUSE	Q9r0g7	mus musculus	480	6	0.4	87	1	MINE_VIBCH	Q9kgn7	vibrio chol
408	7	0.5	1221	1	PBL2_MOUSE	P37889	mus musculus	481	6	0.4	88	1	YB4D_SCHPO	O13973	schizosacch
409	7	0.5	1227	1	B3A3_MOUSE	P16283	mus musculus	482	6	0.4	88	1	YPB4_ECOLI	P03853	escherichia
410	7	0.5	1227	1	B3A3_RAT	P23348	rattus norv	483	6	0.4	89	1	PE23_SHEEP	Q28550	ovis aries
411	7	0.5	1232	1	B3A3_HUMAN	P48751	homo sapien	484	6	0.4	91	1	YOHL_SERPA	P55757	serratia ma
412	7	0.5	1233	1	B3A3_RABIT	O18917	oryctolagus	485	6	0.4	92	1	DYLI_YEAST	Q02647	saccharomyc
413	7	0.5	1233	1	YF16_YEAST	P43597	saccharomyc	486	6	0.4	92	1	H2B_CRONI	P02280	crocodylus
414	7	0.5	1241	1	TRK1_SACBA	P28569	saccharomyc	487	6	0.4	93	1	LE10_HELAN	P46514	heliathanthus
415	7	0.5	1262	1	CVA5_RAT	Q04400	rattus norv	488	6	0.4	93	1	YQIC_CAEEL	Q09283	caenorhabdi
416	7	0.5	1264	1	CVA5_RABIT	P40144	oryctolagus	489	6	0.4	95	1	CXA3_CERLA	P01527	cerebratulu
417	7	0.5	1265	1	CVA5_CANFA	P30803	canis famli	490	6	0.4	96	1	HG14_MOUSE	P18608	mus musculus
418	7	0.5	1361	1	YME9_YEAST	Q04693	saccharomyc	491	6	0.4	96	1	RS18_BORBU	O51140	borrelia bu
419	7	0.5	1377	1	NEO1_RAT	P97603	rattus norv	492	6	0.4	99	1	HMG_TETPY	P40625	tetrahymena
420	7	0.5	1402	1	IF4G_RABIT	P41110	oryctolagus	493	6	0.4	100	1	RS24_METTH	Q26367	methanobact
421	7	0.5	1416	1	BLM_MOUSE	O88700	mus musculus	494	6	0.4	101	1	RS14_CHLPN	Q926w9	chlamydia p
422	7	0.5	1432	1	SK13_YEAST	P17883	saccharomyc	495	6	0.4	101	1	THYP_HUMAN	P20962	homo sapien
423	7	0.5	1443	1	NEO1_CHICK	Q90610	gallus gall	496	6	0.4	103	1	RL21_ECOLI	P04422	escherichia
424	7	0.5	1461	1	NEO1_HUMAN	Q92859	homo sapien	497	6	0.4	103	1	RL21_HAEIN	P44359	haemophilus
425	7	0.5	1493	1	ERC6_HUMAN	Q03468	homo sapien	498	6	0.4	105	1	RL44_YEAST	P02405	saccharomyc
426	7	0.5	1493	1	NEO1_MOUSE	P97798	mus musculus	499	6	0.4	105	1	YJN5_YEAST	P47012	saccharomyc
427	7	0.5	1527	1	CAIH_MOUSE	P39061	mus musculus	500	6	0.4	107	1	YCXQ_PORPU	P51384	porphyra pu
428	7	0.5	1580	1	ACB8_HUMAN	Q09428	homo sapien	501	6	0.4	109	1	LVIF_HUMAN	P04208	homo sapien
429	7	0.5	1581	1	ACC8_CRICR	Q09427	cricetus cr	502	6	0.4	109	1	LVII_HUMAN	P06888	homo sapien
430	7	0.5	1610	1	CCAD_MESAU	Q99244	mesocricetu	503	6	0.4	111	1	LVII_HUMAN	P01699	homo sapien
431	7	0.5	1647	1	SN24_HUMAN	P51532	homo sapien	504	6	0.4	111	1	LVIC_HUMAN	P01701	homo sapien
432	7	0.5	1717	1	RPAL_MOUSE	Q35134	mus musculus	505	6	0.4	111	1	LVID_HUMAN	P01702	homo sapien
433	7	0.5	1720	1	FTSH_CHLYU	P56369	chlorella v	506	6	0.4	112	1	YCX1_CHLRE	P05723	chlamydomon
434	7	0.5	1816	1	KF1B_HUMAN	O60333	homo sapien	507	6	0.4	112	1	LV1B_HUMAN	P01700	homo sapien
435	7	0.5	1816	1	KF1B_MOUSE	Q60575	mus musculus	508	6	0.4	113	1	EXP9_STRPN	P35599	streptococc
436	7	0.5	1820	1	CINA_ELEEL	P02719	electrophor	509	6	0.4	115	1	SOR_PYRHO	O58810	pyrococcus
437	7	0.5	1857	1	FAS2_PENPA	P15368	penicillium	510	6	0.4	116	1	RT11_CHOCR	Q48941	chondrus cr
438	7	0.5	1902	1	SMF1_HUMAN	O14497	homo sapien	511	6	0.4	116	1	TKNK_MOUSE	P55099	mus musculus
439	7	0.5	1966	1	CCAF_HUMAN	O60840	homo sapien	512	6	0.4	116	1	TKNK_RAT	P08435	rattus norv
440	7	0.5	1982	1	CHDM_DROME	O97159	drosophila	513	6	0.4	116	1	XKDD_BACSU	P39783	bacillus su
441	7	0.5	2009	1	SEC7_YEAST	P11075	saccharomyc	514	6	0.4	117	1	VATG_DROME	Q9Xzh6	drosophila
442	7	0.5	2110	1	MCAS_MYCHO	Q02251	mycobacteri	515	6	0.4	120	1	YEO7_SCHPO	O13804	schizosacch
443	7	0.5	2161	1	CCAD_HUMAN	Q01668	homo sapien	516	6	0.4	120	1	YHIT_RICPR	Q9zd11	rickettsia
444	7	0.5	2203	1	CCAD_RAT	P27732	rattus norv	517	6	0.4	121	1	ANFB_MOUSE	P04753	mus musculus
445	7	0.5	2353	1	CCAH_HUMAN	O95180	homo sapien	518	6	0.4	121	1	RBS_ALVHS	P24682	alvinoconch
446	7	0.5	3135	1	S230_PLAFO	Q08372	plasmodium	519	6	0.4	122	1	H2B1_PSAMI	P02287	psammecchinu
447	7	0.5	3418	1	BRC2_HUMAN	P51587	homo sapien	520	6	0.4	122	1	H2B2_PSAMI	P02288	psammecchinu
448	7	0.5	4385	1	YF73_CAEEL	Q09222	caenorhabdi	521	6	0.4	122	1	H2BL_STRPU	P16888	strongyloce
449	7	0.5	4829	1	BIR6_HUMAN	Q9nr09	homo sapien	522	6	0.4	122	1	H2BN_STRPU	P16889	strongyloce
450	7	0.5	5035	1	RYNR_PIG	P16960	sus scrofa	523	6	0.4	122	1	RL7_NEIMA	P80716	neisseria m
451	7	0.5	5037	1	RYNR_RABIT	P11716	oryctolagus	524	6	0.4	123	1	H2BE_STRPU	P02289	strongyloce
452	6	0.4	12	1	UCRH_MOUSE	P99028	mus musculus	525	6	0.4	123	1	H2B_SALTR	Q02282	salmo trutt
453	6	0.4	37	1	RL36_UREPA	Q9pgn7	ureaplasma	526	6	0.4	124	1	H2B_RAT	Q00715	rattus norv
454	6	0.4	47	1	HMC4_DESVH	P33391	desulfovibr	527	6	0.4	124	1	REV_SIVC2	P17280	chimpanzee
455	6	0.4	63	1	CSRA_HAEIN	P44879	haemophilus	528	6	0.4	125	1	H2B1_MOUSE	P10853	mus musculus
456	6	0.4	63	1	ICEL_ASCSU	P07851	ascaris suu	529	6	0.4	125	1	H2B1_XENLA	P02281	xenopus lae
457	6	0.4	64	1	Y592_HAEIN	P44021	haemophilus	530	6	0.4	125	1	H2B2_MOUSE	P10854	mus musculus
458	6	0.4	69	1	YDHZ_ECOLI	P77274	escherichia	531	6	0.4	125	1	H2B2_XENLA	P06900	xenopus lae
459	6	0.4	70	1	RR18_PORPU	P51256	porphyra pu	532	6	0.4	125	1	H2BA_HUMAN	P02278	homo sapien
460	6	0.4	70	1	VC17_SPVKA	P32218	swinepox vi	533	6	0.4	125	1	H2BD_HUMAN	Q99880	homo sapien
461	6	0.4	70	1	Y650_HAEIN	P44028	haemophilus	534	6	0.4	125	1	H2BD_HUMAN	Q99877	homo sapien
462	6	0.4	71	1	RR18_MESVI	Q9mup6	mesostigma	535	6	0.4	125	1	H2BE_HUMAN	Q99879	homo sapien
463	6	0.4	71	1	YVFE_VACCC	P20563	vaccinia vi	536	6	0.4	125	1	H2BF_HUMAN	P33778	homo sapien
464	6	0.4	72	1	MARB_ECOLI	P31121	escherichia	537	6	0.4	125	1	H2BH_HUMAN	Q93078	homo sapien
465	6	0.4	72	1	RL29_CHLTP	P28538	chlamydia t	538	6	0.4	125	1	H2BL_HUMAN	Q93079	homo sapien
466	6	0.4	72	1	RR18_ODOSI	P49505	odontella s	539	6	0.4	125	1	H2BL_HUMAN	Q93080	homo sapien
467	6	0.4	73	1	TDCR_ECOLI	P11866	escherichia	540	6	0.4	125	1	H2BN_HUMAN	P23527	homo sapien
468	6	0.4	73	1	PSBH_MARPO	P12160	marchantia	541	6	0.4	125	1	H2BQ_HUMAN	Q16778	homo sapien
469	6	0.4	75	1	MT1_WHEAT	P43400	trititum ae	542	6	0.4	125	1	H2BR_HUMAN	P06899	homo sapien
470	6	0.4	75	1	PSAE_SYNEN	P25898	synechococcc	543	6	0.4	125	1	H2BS_HUMAN	P57053	homo sapien
471	6	0.4	75	1	RS18_RHOCA	O68127	rhodobacter	544	6	0.4	125	1	H2B_CAIMO	P14001	cairina mos

253	1	527	7	0.5	1	GL9P_HUMAN	PL4314	homo sapien	326	7	0.5	706	1	NUCL_HUMAN	P19338	homo sapien
254	1	528	7	0.5	1	ACH2_CHICK	P09480	gallus gall	327	7	0.5	706	1	NUCL_MOUSE	P09405	mus musculus
255	1	531	7	0.5	1	CALD_RAT	Q62736	rattus norv	328	7	0.5	708	1	ABBI_MOUSE	Q94x11	mus musculus
256	1	538	7	0.5	1	C7C4_MAIZE	Q43257	zea mays (m	329	7	0.5	713	1	NUCL_MSAU	P08199	mesocricetu
257	1	542	7	0.5	1	TULLI_HUMAN	O00294	homo sapien	330	7	0.5	716	1	BAC2_MOUSE	P97303	mus musculus
258	1	544	7	0.5	1	PAK3_HUMAN	O75914	homo sapien	331	7	0.5	721	1	YCF2_OENPI	Q08705	oenothera p
259	1	544	7	0.5	1	PAK3_MOUSE	Q61036	mus musculus	332	7	0.5	728	1	CTCF_CHICK	Q08705	gallus gall
260	1	544	7	0.5	1	PAK3_RAT	Q62829	rattus norv	333	7	0.5	752	1	DRS1_YEAST	P32892	saccharomyc
261	1	545	7	0.5	1	THS_METKA	P50016	methanopyru	334	7	0.5	759	1	SCT1_YEAST	P32784	saccharomyc
262	1	547	7	0.5	1	IF37_MOUSE	O70194	mus musculus	335	7	0.5	763	1	APP2_HUMAN	Q06481	homo sapien
263	1	548	7	0.5	1	IF37_HUMAN	O15371	homo sapien	336	7	0.5	763	1	YN51_YEAST	P42843	saccharomyc
264	1	552	7	0.5	1	YBOO_YEAST	P38222	saccharomyc	337	7	0.5	765	1	APP2_RAT	P15943	rattus norv
265	1	554	7	0.5	1	PEX2_PODAN	P51021	podospora a	338	7	0.5	765	1	UBF1_MOUSE	P25976	mus musculus
266	1	556	7	0.5	1	ECR_MANSE	P49883	manduca sex	339	7	0.5	777	1	RGL2_HUMAN	O15211	homo sapien
267	1	556	7	0.5	1	INX7_CAEEL	Q21123	caenorhabdi	340	7	0.5	780	1	VACL_HUMAN	Q93034	homo sapien
268	1	559	7	0.5	1	PP71_HCMVA	P06726	human cytom	341	7	0.5	780	1	VACL_RABIT	Q29425	oryctolagus
269	1	561	7	0.5	1	SNTC_HUMAN	P49902	homo sapien	342	7	0.5	780	1	VACL_RAT	Q9fj31	rattus norv
270	1	566	7	0.5	1	MXID_SHIFL	Q04641	shigella fl	343	7	0.5	786	1	PRGR_CHICK	P07812	gallus gall
271	1	566	7	0.5	1	MXID_SHISO	Q55293	shigella so	344	7	0.5	792	1	UBPA_YEAST	P53874	saccharomyc
272	1	570	7	0.5	1	NCAP_MOPEI	P19239	mopeia viru	345	7	0.5	795	1	Y21Q_HUMAN	O92609	homo sapien
273	1	575	7	0.5	1	TERM_ADEG1	Q64752	avian adeno	346	7	0.5	799	1	CN3E_MOUSE	O61409	mus musculus
274	1	581	7	0.5	1	DPO4_YEAST	P25615	saccharomyc	347	7	0.5	802	1	NAB3_YEAST	P38996	saccharomyc
275	1	586	7	0.5	1	SYN2_RAT	Q63537	rattus norv	348	7	0.5	809	1	LEF_BACAN	P15917	bacillus an
276	1	586	7	0.5	1	TRM1_ARATH	Q91156	arabidopsis	349	7	0.5	809	1	UPB1_YEAST	P25037	saccharomyc
277	1	587	7	0.5	1	GOX_PENAG	P81156	penicillium	350	7	0.5	817	1	PPSA_PYRFO	P42850	pyrococcus
278	1	587	7	0.5	1	RGPI_HUMAN	P46060	homo sapien	351	7	0.5	819	1	PPSA_PYRAB	Q9v2h7	pyrococcus
279	1	590	7	0.5	1	YM72_YEAST	Q05021	saccharomyc	352	7	0.5	819	1	SWEL_YEAST	P32944	saccharomyc
280	1	591	7	0.5	1	YN48_YEAST	P42846	saccharomyc	353	7	0.5	821	1	PPSA_PYRHO	O57830	pyrococcus
281	1	592	7	0.5	1	LAM2_MOUSE	P21619	mus musculus	354	7	0.5	824	1	ROU_HUMAN	Q00839	homo sapien
282	1	599	7	0.5	1	CENB_HUMAN	P07199	homo sapien	355	7	0.5	825	1	RCAL_YEAST	P40341	saccharomyc
283	1	599	7	0.5	1	CENB_MOUSE	P27790	mus musculus	356	7	0.5	852	1	SRCH_RABIT	P16230	oryctolagus
284	1	601	7	0.5	1	DR1L_MOUSE	Q62431	mus musculus	357	7	0.5	867	1	PROM_MOUSE	O54990	mus musculus
285	1	602	7	0.5	1	PEHX_ERWCH	P15922	erwinia chr	358	7	0.5	872	1	SL45_HUMAN	Q13435	homo sapien
286	1	605	7	0.5	1	GOX_TALFL	Q92452	talaromyces	359	7	0.5	886	1	CN4A_HUMAN	P27815	homo sapien
287	1	606	7	0.5	1	CAC1_YEAST	Q12495	saccharomyc	360	7	0.5	886	1	LEUR_YEAST	P08638	saccharomyc
288	1	606	7	0.5	1	CENB_CRIGR	P48988	cricetulus	361	7	0.5	886	1	NIAL_SOYBN	P54233	glycine max
289	1	608	7	0.5	1	KU7O_MOUSE	P23475	mus musculus	362	7	0.5	886	1	SUHW_DROAN	Q08875	drosophila
290	1	617	7	0.5	1	ABPI_SACEX	P38479	saccharomyc	363	7	0.5	890	1	NIA2_SOYBN	P19870	glycine max
291	1	618	7	0.5	1	SPPA_ECOLI	P08395	escherichia	364	7	0.5	891	1	MAZ3_SCHCO	P37937	schizophyll
292	1	618	7	0.5	1	YKRA_YEAST	P36029	saccharomyc	365	7	0.5	899	1	SUHW_DROVI	Q08876	drosophila
293	1	620	7	0.5	1	GG95_HUMAN	Q08379	homo sapien	366	7	0.5	909	1	CNG4_HUMAN	Q14038	homo sapien
294	1	629	7	0.5	1	YD4B_SCHPO	Q10304	schizosacch	367	7	0.5	912	1	PGCB_BOVIN	Q28062	bos taurus
295	1	630	7	0.5	1	YCF2_OENVI	P31569	oenothera v	368	7	0.5	919	1	PWP2_HUMAN	Q13259	homo sapien
296	1	635	7	0.5	1	SL56_HUMAN	Q9y289	homo sapien	369	7	0.5	924	1	KEX2_CANAL	O13359	candida alb
297	1	636	7	0.5	1	NTPR_HUMAN	Q99884	homo sapien	370	7	0.5	929	1	RBMA_HUMAN	P98175	homo sapien
298	1	644	7	0.5	1	NFM_RABIT	P54938	oryctolagus	371	7	0.5	932	1	YATA_SCHPO	Q09897	schizosacch
299	1	645	7	0.5	1	PDA4_HUMAN	P13667	homo sapien	372	7	0.5	960	1	FGDI_MOUSE	P52734	mus musculus
300	1	646	7	0.5	1	WEEL_HUMAN	P30291	homo sapien	373	7	0.5	960	1	VP41_LYCES	P93231	lycopersico
301	1	646	7	0.5	1	WEEL_MOUSE	P47810	mus musculus	374	7	0.5	960	1	Y682_HUMAN	Q9y4c8	homo sapien
302	1	652	7	0.5	1	DREB_CHICK	P18302	gallus gall	375	7	0.5	963	1	REF1_MOUSE	P48377	mus musculus
303	1	653	7	0.5	1	C1K4_HUMAN	P22459	homo sapien	376	7	0.5	967	1	HEX_ADE02	P03277	human adeno
304	1	654	7	0.5	1	C1K4_MOUSE	Q61423	mus musculus	377	7	0.5	971	1	NAM7_YEAST	P30771	saccharomyc
305	1	654	7	0.5	1	C1K4_MOUSE	Q28527	mus musculus	378	7	0.5	975	1	KINH_DROME	P17210	drosophila
306	1	655	7	0.5	1	C1K4_MOUSE	P15385	rattus norv	379	7	0.5	976	1	VP41_ARATH	P93043	arabidopsis
307	1	661	7	0.5	1	NTPR_RAT	P28573	rattus norv	380	7	0.5	979	1	REF1_HUMAN	P22670	homo sapien
308	1	663	7	0.5	1	TERM_ADEB3	O55439	bovine aden	381	7	0.5	999	1	OXRP_CRIGR	Q60432	cricetulus
309	1	665	7	0.5	1	FILS_HUMAN	Q12934	homo sapien	382	7	0.5	999	1	OXRE_RAT	Q63617	rattus norv
310	1	669	7	0.5	1	SSRP_CHICK	Q04678	gallus gall	383	7	0.5	1034	1	CAPP_SYNY3	P74299	synecocyst
311	1	674	7	0.5	1	NUOL_NEIMA	Q9jx92	neisseria m	384	7	0.5	1041	1	DD16_HUMAN	O60231	homo sapien
312	1	674	7	0.5	1	NUOL_NEIMB	Q9k1b0	neisseria m	385	7	0.5	1043	1	TCF8_MESAU	Q60542	mesocricetu
313	1	675	7	0.5	1	VPS5_YEAST	Q92331	saccharomyc	386	7	0.5	1058	1	PMAL_DICDI	P54679	dictyosteli
314	1	678	7	0.5	1	C1CL_RABIT	P51804	oryctolagus	387	7	0.5	1062	1	SUM1_YEAST	P46676	saccharomyc
315	1	678	7	0.5	1	GARP_PLAFF	P13816	plasmodium	388	7	0.5	1067	1	SGG_DROME	P18431	drosophila
316	1	679	7	0.5	1	YHCS_YEAST	P38738	saccharomyc	389	7	0.5	1078	1	CYA7_BOVIN	Q29450	bos taurus
317	1	684	7	0.5	1	RPSD_RHIME	Q59753	rhizobium m	390	7	0.5	1080	1	CYA7_HUMAN	P51828	homo sapien
318	1	687	7	0.5	1	C1CK_HUMAN	P51800	homo sapien	391	7	0.5	1099	1	CYA7_MOUSE	P51829	mus musculus
319	1	687	7	0.5	1	C1CK_RABIT	P51803	oryctolagus	392	7	0.5	1108	1	CN3B_RAT	Q63085	rattus norv
320	1	687	7	0.5	1	C1CK_RAT	Q06393	rattus norv	393	7	0.5	1109	1	TCF8_RAT	Q62947	rattus norv
321	1	687	7	0.5	1	C1CL_RAT	P51802	rattus norv	394	7	0.5	1113	1	HDA3_MOUSE	Q922v6	mus musculus
322	1	692	7	0.5	1	YK06_YEAST	P36062	saccharomyc	395	7	0.5	1114	1	TCF8_CHICK	P36197	gallus gall
323	1	695	7	0.5	1	APP2_MOUSE	Q06335	mus musculus	396	7	0.5	1117	1	TCF8_MOUSE	Q64318	mus musculus
324	1	699	7	0.5	1	SRCH_HUMAN	P23327	homo sapien	397	7	0.5	1123	1	RBME_HUMAN	P78332	homo sapien
325	1	704	7	0.5	1	HPS_MOUSE	O08983	mus musculus	398	7	0.5	1124	1	TCF8_HUMAN	P37275	homo sapien

107	1	267	7	0.5	180	7	0.5	406	1	GCST_SOLTU	P54260 solanum tub
108	1	268	7	0.5	181	7	0.5	406	1	MYC_BRARE	P52160 brachydanio
109	1	272	7	0.5	182	7	0.5	411	1	FKB3_YEAST	P38911 saccharomyc
110	1	273	7	0.5	183	7	0.5	411	1	MP62_LYTP1	P57753 lytechinus
111	1	278	7	0.5	184	7	0.5	412	1	FW44_HUMAN	P51775 homo sapien
112	1	278	7	0.5	185	7	0.5	412	1	Y087_YEAST	Q26486 spodoptera
113	1	279	7	0.5	186	7	0.5	416	1	ASF1_YEAST	Q28173 bos taurus
114	1	280	7	0.5	187	7	0.5	423	1	Y326_THEMA	Q08168 plasmodium
115	1	282	7	0.5	188	7	0.5	427	1	YKRG_ECOLI	P36049 saccharomyc
116	1	288	7	0.5	189	7	0.5	430	1	SSRA_ONCMY	P36916 mus musculu
117	1	291	7	0.5	190	7	0.5	430	1	CAH8_MOUSE	P36916 mus musculu
118	1	292	7	0.5	191	7	0.5	433	1	Y152_HUMAN	P35205 rattus norv
119	1	293	7	0.5	192	7	0.5	435	1	RY44_SPIO1	P30283 saccharomyc
120	1	295	7	0.5	193	7	0.5	439	1	CGDI_HUMAN	Q46107 drosophila
121	1	295	7	0.5	194	7	0.5	440	1	YDT5_SCHPO	Q98937 gallus gall
122	1	299	7	0.5	195	7	0.5	445	1	YDT5_SCHPO	Q98937 gallus gall
123	1	299	7	0.5	196	7	0.5	445	1	YDT5_SCHPO	Q98937 gallus gall
124	1	300	7	0.5	197	7	0.5	446	1	YDT5_SCHPO	Q98937 gallus gall
125	1	302	7	0.5	198	7	0.5	446	1	YDT5_SCHPO	Q98937 gallus gall
126	1	303	7	0.5	199	7	0.5	448	1	YDT5_SCHPO	Q98937 gallus gall
127	1	304	7	0.5	200	7	0.5	448	1	YDT5_SCHPO	Q98937 gallus gall
128	1	309	7	0.5	201	7	0.5	449	1	YDT5_SCHPO	Q98937 gallus gall
129	1	310	7	0.5	202	7	0.5	449	1	YDT5_SCHPO	Q98937 gallus gall
130	1	312	7	0.5	203	7	0.5	450	1	YDT5_SCHPO	Q98937 gallus gall
131	1	315	7	0.5	204	7	0.5	452	1	YDT5_SCHPO	Q98937 gallus gall
132	1	316	7	0.5	205	7	0.5	452	1	YDT5_SCHPO	Q98937 gallus gall
133	1	317	7	0.5	206	7	0.5	454	1	YDT5_SCHPO	Q98937 gallus gall
134	1	317	7	0.5	207	7	0.5	455	1	YDT5_SCHPO	Q98937 gallus gall
135	1	320	7	0.5	208	7	0.5	457	1	YDT5_SCHPO	Q98937 gallus gall
136	1	320	7	0.5	209	7	0.5	460	1	YDT5_SCHPO	Q98937 gallus gall
137	1	324	7	0.5	210	7	0.5	462	1	YDT5_SCHPO	Q98937 gallus gall
138	1	328	7	0.5	211	7	0.5	463	1	YDT5_SCHPO	Q98937 gallus gall
139	1	329	7	0.5	212	7	0.5	464	1	YDT5_SCHPO	Q98937 gallus gall
140	1	329	7	0.5	213	7	0.5	467	1	YDT5_SCHPO	Q98937 gallus gall
141	1	329	7	0.5	214	7	0.5	467	1	YDT5_SCHPO	Q98937 gallus gall
142	1	329	7	0.5	215	7	0.5	470	1	YDT5_SCHPO	Q98937 gallus gall
143	1	329	7	0.5	216	7	0.5	471	1	YDT5_SCHPO	Q98937 gallus gall
144	1	330	7	0.5	217	7	0.5	471	1	YDT5_SCHPO	Q98937 gallus gall
145	1	331	7	0.5	218	7	0.5	474	1	YDT5_SCHPO	Q98937 gallus gall
146	1	333	7	0.5	219	7	0.5	474	1	YDT5_SCHPO	Q98937 gallus gall
147	1	335	7	0.5	220	7	0.5	475	1	YDT5_SCHPO	Q98937 gallus gall
148	1	336	7	0.5	221	7	0.5	477	1	YDT5_SCHPO	Q98937 gallus gall
149	1	336	7	0.5	222	7	0.5	477	1	YDT5_SCHPO	Q98937 gallus gall
150	1	336	7	0.5	223	7	0.5	478	1	YDT5_SCHPO	Q98937 gallus gall
151	1	336	7	0.5	224	7	0.5	480	1	YDT5_SCHPO	Q98937 gallus gall
152	1	338	7	0.5	225	7	0.5	481	1	YDT5_SCHPO	Q98937 gallus gall
153	1	340	7	0.5	226	7	0.5	482	1	YDT5_SCHPO	Q98937 gallus gall
154	1	344	7	0.5	227	7	0.5	482	1	YDT5_SCHPO	Q98937 gallus gall
155	1	344	7	0.5	228	7	0.5	482	1	YDT5_SCHPO	Q98937 gallus gall
156	1	350	7	0.5	229	7	0.5	482	1	YDT5_SCHPO	Q98937 gallus gall
157	1	356	7	0.5	230	7	0.5	484	1	YDT5_SCHPO	Q98937 gallus gall
158	1	357	7	0.5	231	7	0.5	485	1	YDT5_SCHPO	Q98937 gallus gall
159	1	357	7	0.5	232	7	0.5	485	1	YDT5_SCHPO	Q98937 gallus gall
160	1	357	7	0.5	233	7	0.5	491	1	YDT5_SCHPO	Q98937 gallus gall
161	1	368	7	0.5	234	7	0.5	493	1	YDT5_SCHPO	Q98937 gallus gall
162	1	368	7	0.5	235	7	0.5	495	1	YDT5_SCHPO	Q98937 gallus gall
163	1	371	7	0.5	236	7	0.5	497	1	YDT5_SCHPO	Q98937 gallus gall
164	1	375	7	0.5	237	7	0.5	502	1	YDT5_SCHPO	Q98937 gallus gall
165	1	375	7	0.5	238	7	0.5	502	1	YDT5_SCHPO	Q98937 gallus gall
166	1	382	7	0.5	239	7	0.5	502	1	YDT5_SCHPO	Q98937 gallus gall
167	1	394	7	0.5	240	7	0.5	503	1	YDT5_SCHPO	Q98937 gallus gall
168	1	382	7	0.5	241	7	0.5	504	1	YDT5_SCHPO	Q98937 gallus gall
169	1	382	7	0.5	242	7	0.5	504	1	YDT5_SCHPO	Q98937 gallus gall
170	1	383	7	0.5	243	7	0.5	504	1	YDT5_SCHPO	Q98937 gallus gall
171	1	383	7	0.5	244	7	0.5	505	1	YDT5_SCHPO	Q98937 gallus gall
172	1	392	7	0.5	245	7	0.5	507	1	YDT5_SCHPO	Q98937 gallus gall
173	1	394	7	0.5	246	7	0.5	511	1	YDT5_SCHPO	Q98937 gallus gall
174	1	394	7	0.5	247	7	0.5	515	1	YDT5_SCHPO	Q98937 gallus gall
175	1	396	7	0.5	248	7	0.5	516	1	YDT5_SCHPO	Q98937 gallus gall
176	1	396	7	0.5	249	7	0.5	517	1	YDT5_SCHPO	Q98937 gallus gall
177	1	399	7	0.5	250	7	0.5	518	1	YDT5_SCHPO	Q98937 gallus gall
178	1	400	7	0.5	251	7	0.5	521	1	YDT5_SCHPO	Q98937 gallus gall
179	1	401	7	0.5	252	7	0.5	522	1	YDT5_SCHPO	Q98937 gallus gall

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:05:29 ; Search time 16.55 seconds
(without alignments)
2879.117 Million cell updates/sec

Title: US-09-512-581-2
Perfect score: 1391
Sequence: 1 MAHSTRINDGKITPPGVK.....QKGRGPKSTPSPQPKNV 1391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.6	439	1 TRF1_HUMAN	P54274 homo sapien
2	9	0.6	1938	1 MYSA_MOUSE	Q02566 mus musculus
3	9	0.6	1938	1 MYSA_MOUSE	Q02563 rattus norv
4	8	0.6	87	1 GBG_LOLFO	Q01821 loligo forb
5	8	0.6	91	1 UCRH_HUMAN	P07919 homo sapien
6	8	0.6	181	1 VE4_HPV04	Q07852 human papill
7	8	0.6	206	1 HMG2_CHICK	P26584 gallus gall
8	8	0.6	233	1 EMB1_CAVPO	P22032 cavia porce
9	8	0.6	233	1 SOML_ONCKE	P24405 oncorhynch
10	8	0.6	242	1 THYL_HUMAN	P20396 homo sapien
11	8	0.6	316	1 YNFR_YEAST	P53947 saccharomyc
12	8	0.6	339	1 RLAO_ARCFU	Q28781 archaeoglob
13	8	0.6	418	1 YSE2_CAEEL	Q09936 caenorhabdi
14	8	0.6	451	1 GP1D_CHLTR	P10555 chlamydia t
15	8	0.6	463	1 CMGA_MOUSE	P26339 mus musculus
16	8	0.6	466	1 CMGA_MOUSE	P10354 rattus norv
17	8	0.6	513	1 NIEK_RHLSN	P19067 rhizobium s
18	8	0.6	533	1 LCP2_HUMAN	Q13094 homo sapien
19	8	0.6	544	1 INVO_AOTTR	P24708 aotus trivi
20	8	0.6	544	1 NFL_XENLA	P35616 xenopus lae
21	8	0.6	585	1 LAMI_HUMAN	P20700 homo sapien
22	8	0.6	587	1 LAMI_MOUSE	P14733 mus musculus
23	8	0.6	653	1 APPL_MOUSE	Q03157 mus musculus
24	8	0.6	823	1 UBPG_HUMAN	Q95455 homo sapien
25	8	0.6	1002	1 IF2P_YEAST	P39730 saccharomyc
26	8	0.6	1081	1 SPS2_CRAPL	Q04933 craterostig
27	8	0.6	1165	1 CYA6_CANFA	P30804 canis fami
28	8	0.6	1165	1 CYA6_MOUSE	Q01341 mus musculus
29	8	0.6	1166	1 CYA6_RAT	Q03343 rattus norv
30	8	0.6	1168	1 CYA6_HUMAN	Q43306 homo sapien
31	8	0.6	1184	1 FBL2_HUMAN	P98095 homo sapien
32	8	0.6	1220	1 IF2P_HUMAN	Q60841 homo sapien
33	8	0.6	1859	1 GBF1_HUMAN	Q92538 homo sapien

1	1934	0.6	8	34	1	MYSB_MESAU	P13540 mesocricetu
1	1935	0.6	8	35	1	MYSB_RAT	P02564 rattus norv
1	2248	0.6	8	36	1	CYAL_DROME	P32870 drosophila
1	2375	0.6	8	37	1	ATRA_HUMAN	P46100 homo sapien
1	73	0.5	7	38	1	RR18_GUTH	O78488 guillardia
1	78	0.5	7	39	1	UCRH_BOVIN	P00126 bos taurus
1	96	0.5	7	40	1	E111_ADEM1	P12533 mouse adeno
1	97	0.5	7	41	1	VE7_COPV	O89759 canine oral
1	98	0.5	7	42	1	SNCS_HUMAN	O75971 homo sapien
1	101	0.5	7	43	1	RLI2_METTH	P05394 methanobact
1	101	0.5	7	44	1	THYP_BOVIN	P08814 bos taurus
1	101	0.5	7	45	1	THYP_RAT	P04550 rattus norv
1	102	0.5	7	46	1	HSP3_MOUSE	O62100 mus musculu
1	105	0.5	7	47	1	HXA7_RAT	P09634 rattus norv
1	106	0.5	7	48	1	RLI2_ARCFU	O28780 archaeoglob
1	109	0.5	7	49	1	THYA_BOVIN	P01252 bos taurus
1	110	0.5	7	50	1	THYA_HUMAN	P06454 homo sapien
1	110	0.5	7	51	1	THYA_MOUSE	P26350 mus musculu
1	111	0.5	7	52	1	THVA_RAT	P06302 rattus norv
1	112	0.5	7	53	1	RLA2_PLAFA	O00806 plasmodium
1	116	0.5	7	54	1	GCAD_BACCL	P42817 bacillus ca
1	120	0.5	7	55	1	SMS2_CARAU	O9ygn4 carassius a
1	127	0.5	7	56	1	RBFA_BORBU	O51742 borrelia bu
1	133	0.5	7	57	1	RS19_ARCFU	O28358 archaeoglob
1	134	0.5	7	58	1	CLX2_HUMAN	Q13329 homo sapien
1	137	0.5	7	59	1	IPPD_PIG	O29277 sus scrofa
1	142	0.5	7	60	1	YN8P_YEAST	P53737 saccharomyc
1	152	0.5	7	61	1	HMBE_SOYBN	Q10370 glycine max
1	154	0.5	7	62	1	PUTR_RHOCA	O52710 rhodobacter
1	167	0.5	7	63	1	E1BS_ADE40	P10543 human adeno
1	170	0.5	7	64	1	E1BS_ADE41	P10544 human adeno
1	173	0.5	7	65	1	SODC_ECOLI	P53635 escherichia
1	176	0.5	7	66	1	HMG2_SOYBN	Q00423 glycine max
1	180	0.5	7	67	1	HMG1_CRIGR	P07136 cricetus
1	181	0.5	7	68	1	E320_ADE1A	P35767 human adeno
1	181	0.5	7	69	1	E320_ADE1P	P35768 human adeno
1	181	0.5	7	70	1	E320_ADE35	P15137 human adeno
1	182	0.5	7	71	1	Y314_ARATH	O9m8t3 arabidopsis
1	184	0.5	7	72	1	UBC4_WHEAT	P16577 triticum ae
1	200	0.5	7	73	1	NUPL_XENLA	P05221 xenopus lae
1	202	0.5	7	74	1	IPPD_BOVIN	P07516 bos taurus
1	204	0.5	7	75	1	IPPD_HUMAN	O9ud71 homo sapien
1	206	0.5	7	76	1	6B_AGRVI	O04551 agrobacteri
1	206	0.5	7	77	1	RHTC_ECOLI	P27846 escherichia
1	206	0.5	7	78	1	S3AE_BACSU	P49783 bacillus su
1	208	0.5	7	79	1	HMG2_HUMAN	P26583 homo sapien
1	209	0.5	7	80	1	HMG2_MOUSE	P30681 mus musculu
1	209	0.5	7	81	1	HMG2_PIG	P17741 sus scrofa
1	209	0.5	7	82	1	HMG2_RAT	P52925 rattus norv
1	209	0.5	7	83	1	VDGI_BACSU	P96707 bacillus su
1	211	0.5	7	84	1	RGS2_HUMAN	P41220 homo sapien
1	213	0.5	7	85	1	NEUM_CARAU	P17691 carassius a
1	213	0.5	7	86	1	VE4_HPV65	Q07873 human papil
1	214	0.5	7	87	1	HMG1_MOUSE	P07155 mus musculu
1	215	0.5	7	88	1	CALL_ARATH	P30188 arabidopsis
1	220	0.5	7	89	1	HB7B_XENLA	P04476 xenopus lae
1	224	0.5	7	90	1	TUR8_MOUSE	P19473 mus musculu
1	225	0.5	7	91	1	HT14_ARATH	P46685 arabidopsis
1	227	0.5	7	92	1	EMBP_RAT	O63189 rattus norv
1	229	0.5	7	93	1	HXA7_MOUSE	P02830 mus musculu
1	233	0.5	7	94	1	FSA_XENLA	P31515 xenopus lae
1	234	0.5	7	95	1	EMB2_CAVPO	P35709 cavia porce
1	238	0.5	7	96	1	VGL6_HSV11	P06484 herpes simp
1	239	0.5	7	97	1	CENB_SHEEP	P49451 ovis aries
1	247	0.5	7	98	1	TRPD_BACCA	P30525 bacillus ca
1	247	0.5	7	99	1	EFTS_SPIPL	P34828 spirulina p
1	248	0.5	7	100	1	MTRA_METKA	O32867 methanopyru
1	249	0.5	7	101	1	UL07_HSV6Z	P52457 human herpe
1	256	0.5	7	102	1	YKJ9_YEAST	P34247 saccharomyc
1	261	0.5	7	103	1	RS3A_ORISA	P49397 oryza sativ
1	261	0.5	7	104	1	VP31_FRG3V	P18178 frog virus
1	263	0.5	7	105	1	TRPA_RHOSH	O9x468 rhodobacter
1	265	0.5	7	106	1	EN16_STRPU	P13665 strongyloce

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 307;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 74 eeeeeer 81

Search completed: September 25, 2001, 10:06:18
Job time: 184 sec

RESULT 50
AAG37417
ID AAG37417 standard; Protein; 307 AA.
XX
AC AAG37417;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46004.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 17-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 307;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEEEER 1332
DB 7A eeeeeer 81

05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.
07-SEP-1999; 99US-0152363.
10-SEP-1999; 99US-0153070.
13-SEP-1999; 99US-0153758.
15-SEP-1999; 99US-0154018.
16-SEP-1999; 99US-0154039.
20-SEP-1999; 99US-0154779.
22-SEP-1999; 99US-0155139.
23-SEP-1999; 99US-0155486.
24-SEP-1999; 99US-0155659.
28-SEP-1999; 99US-0156458.
29-SEP-1999; 99US-0156596.
04-OCT-1999; 99US-0157117.
05-OCT-1999; 99US-0157753.
06-OCT-1999; 99US-0157865.
07-OCT-1999; 99US-0158029.
08-OCT-1999; 99US-0158232.
12-OCT-1999; 99US-0158369.
13-OCT-1999; 99US-0159293.
13-OCT-1999; 99US-0159294.
13-OCT-1999; 99US-0159295.
14-OCT-1999; 99US-0159329.
14-OCT-1999; 99US-0159330.
14-OCT-1999; 99US-0159331.
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0159638.
18-OCT-1999; 99US-0159584.
21-OCT-1999; 99US-0160741.
21-OCT-1999; 99US-0160767.
21-OCT-1999; 99US-0160768.
21-OCT-1999; 99US-0160770.
21-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160980.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.
25-OCT-1999; 99US-0161404.
25-OCT-1999; 99US-0161405.
25-OCT-1999; 99US-0161406.
26-OCT-1999; 99US-0161359.
26-OCT-1999; 99US-0161360.
26-OCT-1999; 99US-0161361.
28-OCT-1999; 99US-0161920.
28-OCT-1999; 99US-0161992.
28-OCT-1999; 99US-0161993.
29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 294;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
Db 226 peeeeeee 233
|||||||
RESULT 49
AAG16481
ID AAG16481 standard; Protein; 307 AA.
XX
AC AAG16481;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17144.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 260;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

OY 1325 REEEERER 1332
Db 27 eeeeeee 34

RESULT 48

AAAG31669
ID AAG31669 standard; Protein; 294 AA.

XX AAG31669;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38071.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147303.

PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.


```
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      0.6%; Score 8; DB 21; Length 260;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db 27 eeeeeer 34
|||||||

RESULT 47
AAG38888
ID AAG38888 standard; Protein; 260 AA.
XX
AC AAG38888;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48041.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
```


PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 0.6%; Score 8; DB 21; Length 254;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 21 eeeeeer 28

RESULT 46

AAG08527

ID AAG08527 standard; Protein; 260 AA.

XX AC AAG08527;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 6099.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

DR WPI; 2001-182940/18.
 XX N-PSDB; AAF30409.
 PT New isolated or recombinant polynucleotide encoding an antigenic
 PT polypeptide (cytokine IL-D80) for providing reagents to regulate
 PT activation, development, differentiation and function of cells
 PT including haematopoietic cells -
 XX
 XX
 XX Claim 1; Page 9; 73pp; English.
 XX The present sequence is that of a novel human cytokine, termed
 CC variant interleukin DNAX 80 (IL-D80), which shows homology
 CC to interleukin-11. The sequence was deduced from isolated IL-D80
 CC cDNA (see AAF30409). A human IL-D80 sequence is also provided in
 CC AAB20275. The 2 IL-D80 sequences differ only in the signal peptide
 CC region. IL-D80 probably has either stimulatory or inhibitory
 CC effects on haematopoietic cells, including lymphoid cells such as
 CC T-cells, B-cells, natural killer cells, macrophages, dendritic
 CC cells, haematopoietic progenitors, etc. It may be useful in the
 CC treatment of immune disorders, such as T-cell immune deficiency,
 CC chronic inflammation or tissue rejection, or in cardiovascular or
 CC neurophysiological conditions. The full-length cytokine, and its
 CC fragments or antagonists, will be useful in physiological modulation
 CC of cells expressing a receptor. The proteins will also be useful as
 CC antigens, e.g. immunogens, for raising antibodies to various
 CC epitopes on the protein. Human and murine IL-D80 polypeptides and
 CC polynucleotides are claimed, as well as expression vectors, host
 CC cells, antigenic polypeptides, a detection kit, and a method of
 CC modulating the physiology or development of a cell or tissue
 CC culture cell by contacting with an agonist or antagonist of
 CC primate IL-D80.
 XX
 XX Sequence 243 AA;
 SQ
 Query Match 0.6%; Score 8; DB 22; Length 243;
 Best Local Similarity 100.08; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1324 PEEEEEE 1331
 Db 163 peeeeeee 170
 |||||
 RESULT 44
 AAG16482
 ID AAG16482 standard; Protein; 254 AA.
 XX
 AC AAG16482;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 17145.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 PF
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 04-MAY-1999; 99US-0132048.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 07-MAY-1999; 99US-0132486.
 PR 11-MAY-1999; 99US-0132487.
 PR 14-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 24-MAY-1999; 99US-0135353.
 PR 25-MAY-1999; 99US-0135629.
 PR 27-MAY-1999; 99US-0136021.
 PR 28-MAY-1999; 99US-0136392.
 PR 01-JUN-1999; 99US-0136782.
 PR 03-JUN-1999; 99US-0137222.
 PR 04-JUN-1999; 99US-0137528.
 PR 07-JUN-1999; 99US-0137502.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139453.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.

Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||

Db 169 eeeeeee 176

RESULT 42

AAB20275
ID AAB20275 standard; Protein; 242 AA.

XX AC AAB20275;

XX DT 14-MAY-2001 (first entry)

XX DE Human interleukin DNAX 80.

XX KW Interleukin DNAX 80; IL-D80; human; cytokine; immunomodulator.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..27

FT /label= Signal_peptide

FT Protein

FT /label= Mature_protein

FT Region

FT /label= Helix-A

FT Region

FT /label= Helix-B

FT Region

FT /label= Helix-C

FT Region

FT /label= Helix-D

FT Region

FT /note= "leucine repeat region"

FT Region

FT /note= "glutamic acid repeat region"

FT Region

FT /note= "leucine repeat region"

XX WO200109176-A2.

XX PD 08-FEB-2001.

XX PF 27-JUL-2000; 2000WO-US20475.

XX PR 30-JUL-1999; 99US-0364674.

XX PR 06-AUG-1999; 99US-0369643.

XX XX (SCHE) SCHERING CORP.

XX PA Timans JC, Kastelein RA, Bazan JF;

XX PI WPI: 2001-182940/18.

XX DR N-PSDB; AAF30407.

XX DR New isolated or recombinant polynucleotide encoding an antigenic

XX PT polypeptide (cytokine IL-D80) for providing reagents to regulate

XX PT activation, development, differentiation and function of cells

XX PT including haematopoietic cells

XX PS Claim 1; Page 8; 73pp; English.

XX XX The present sequence is that of a novel human cytokine, termed

XX CC interleukin DNAX 80, or IL-D80, which shows sequence homology to

XX CC interleukin-11. The sequence was deduced from isolated IL-D80

XX CC cDNA (see AAF30407). A variant huIL-D80 sequence is provided in

XX CC AAB20277. The 2 sequences differ only in the signal peptide

XX CC region. IL-D80 probably has either stimulatory or inhibitory

XX CC effects on haematopoietic cells, including lymphoid cells such as

XX CC T-cells, B-cells, natural killer cells, macrophages, dendritic

CC cells, haematopoietic progenitors, etc. It may be useful in the
CC treatment of immune disorders, such as T-cell immune deficiency,
CC chronic inflammation or tissue rejection, or in cardiovascular or
CC neurophysiological conditions. The full-length cytokine, and its
CC fragments or antagonists, will be useful in physiological modulation
CC of cells expressing a receptor. The proteins will also be useful as
CC antigens, e.g. immunogens, for raising antibodies to various
CC epitopes on the protein. Human and murine IL-D80 polypeptides and
CC polynucleotides are claimed, as well as expression vectors, host
CC cells, antigenic polypeptides, a detection kit, and a method of
CC modulating the physiology or development of a cell or tissue
CC culture cell by contacting with an agonist or antagonist of
CC primate IL-D80.

XX XX Sequence 242 AA;

Query Match 0.6%; Score 8; DB 22; Length 242;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

Db 162 peeeeeee 169

RESULT 43

AAB20277
ID AAB20277 standard; Protein; 243 AA.

XX AC AAB20277;

XX DT 14-MAY-2001 (first entry)

XX DE Human interleukin DNAX 80 variant.

XX KW Interleukin DNAX 80; IL-D80; human; cytokine; immunomodulator.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..28

FT /label= Signal_peptide

FT Protein 29..243

FT /label= Mature_protein

FT Region 35..57

FT /label= Helix-A

FT Region 89..114

FT /label= Helix-B

FT Region 124..157

FT /label= Helix-C

FT Region 204..231

FT /label= Helix-D

FT Region 14..21

FT /note= "leucine repeat region"

FT Region 164..176

FT /note= "glutamic acid repeat region"

FT Region 221..224

FT /note= "leucine repeat region"

XX WO200109176-A2.

XX PN 08-FEB-2001.

XX XX 27-JUL-2000; 2000WO-US20475.

XX PR 30-JUL-1999; 99US-0364674.

XX PR 06-AUG-1999; 99US-0369643.

XX XX (SCHE) SCHERING CORP.

XX PA Timans JC, Kastelein RA, Bazan JF;

XX PI Timans JC, Kastelein RA, Bazan JF;

PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 237;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 EEEEEER 1332
 DB 4 eeeeeer 11

RESULT 40
 AAR31229
 ID AAR31229 standard; protein; 242 AA.
 XX
 AC AAR31229;

DT 19-MAY-1993 (first entry)
 DE Prepro-thyrotropin releasing hormone.
 XX
 KW Prepro TRH; antibodies; human; pituitary dysfunction; psychoses;
 KW thyroidal diseases; thyroid.

OS Homo sapiens.
 XX
 PN JP04352797-A.
 XX
 PD 07-DEC-1992.
 XX
 PF 29-MAY-1991; 91JP-0228139.
 XX
 PR 29-MAY-1991; 91JP-0228139.
 XX
 PA (SAKA) OTSUKA PHARM CO LTD.
 XX
 DR WPI; 1993-024099/03.

PT Human prepro-TRH related peptide(s) - used to prepare antibodies
 PT which react specifically with human prepro-TRH for diagnosis of

PT e.g. pituitary dysfunction etc.
 XX
 PS Disclosure; Page 14; 19pp; Japanese.
 XX
 CC The sequence shown is that of human prepro-thyrotropin releasing
 CC hormone (TRH). The protein may be used to prepare related peptides
 CC which can be used to raise antibodies which specifically react with
 CC human prepro-TRH. The resultant antibodies may be used for the
 CC diagnosis of human TRH related diseases, e.g. pituitary dysfunction,
 CC thyroidal diseases and psychoses.
 CC See also AAR31230-6.
 XX
 SQ Sequence 242 AA;

Query Match 0.6%; Score 8; DB 14; Length 242;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 EEEEEER 1332
 DB 169 eeeeeer 176

RESULT 41
 AAW68538
 ID AAW68538 standard; Protein; 242 AA.

XX
 AC AAW68538;

DT 12-FEB-1999 (first entry)

DE Pepto TRH protein, a TRAP molecule.

KW Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;
 KW lymphoma; cancer; HLA; human lymphocyte antigen; vaccine.

OS Homo sapiens.

PN WO9849299-A1.

PD 05-NOV-1998.

PF 22-APR-1998; 98WO-US07784.

PR 25-APR-1997; 97US-0845998.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Boon-Falleur T, Coulie PG, De Smet C, Lucas S, Van Baren N;

DR WPI; 1999-009425/01.

DR N-PSDB; AAV33909.

PT New diagnosis of leukaemia - by detecting genes for tumour antigen
 PT rejection precursors or corresponding proteins

PS Claim 11; Pages 54-55; 88pp; English.

CC The present sequence represents a tumour rejection antigen precursor
 CC (TRAP) designated TRH. The specification describes the treatment of
 CC disorders which characterised by expression of a leukaemia-associated
 CC nucleic acid such as TRH. The products are used for in vivo or in vitro
 CC screening for leukaemia, lymphoma or other cancers by usual
 CC hybridisation/amplification or immunoassay methods. TRAPS when
 CC processed to antigens or complexed with HLA (human lymphocyte antigen)
 CC molecules, or nucleic acid encoding them, are useful in vaccines for
 CC treating leukaemia.

XX Sequence 242 AA;

Query Match 0.6%; Score 8; DB 20; Length 242;

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149502.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 237;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db |||||
4 eeeeeer 11

RESULT 39

AAG37419
ID AAG37419 standard; Protein; 237 AA.

XX
AC AAG37419;

XX
DT 18-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46006.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX


```

AC AAW40075;
XX
XX 29-MAY-1998 (first entry)
XX
XX Guinea pig eosinophil granule MBP GMBP-1.
XX
XX Eosinophil-derived basic protein; EBPH; diagnosis; prevention; treatment;
XX disease; eosinophil accumulation; granule release; allergic reaction;
XX inflammatory reaction; eosinophilias; parasitic infection; pregnancy;
XX cytolytic agent; bacterial infection; cancer; major basic protein;
XX MBP; guinea pig.
XX
XX Cavia sp.
XX
XX US5728820-A.
XX
XX 17-MAR-1998.
XX
XX 23-OCT-1996; 96US-0740036.
XX
XX 23-OCT-1996; 96US-0740036.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Akerblom IE;
XX
XX WPI; 1998-206622/18.
XX
XX New DNA encoding eosinophil-derived basic protein - useful for
XX producing recombinant protein used for treating, e.g. parasitic
XX infections
XX
XX Disclosure; Fig 2; 26pp; English.
XX
XX This sequence represents the guinea pig eosinophil granule major basic
XX protein (MBP), GMBP-1, which is used to identify an eosinophil-derived
XX basic protein (EBPH) derived from interleukin-5 (IL-5) cultured human
XX umbilical cord blood cells. This protein can be used in the diagnosis,
XX prevention and treatment of diseases and conditions associated with
XX eosinophil accumulation and granule release, including late phase
XX allergic/inflammatory reactions, eosinophilias, parasitic infections and
XX conditions associated with placental-derived, eosinophil-derived basic
XX protein accumulation in pregnancy. The protein can also be used for
XX producing recombinant EBPH, which may be useful as a cytolytic agent in
XX the treatment of e.g. bacterial and parasitic infections and cancer.
XX
XX Sequence 233 AA;
XX

Query Match 0.6%; Score 8; DB 19; Length 233;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
Db 73 peeeeeee 80

RESULT 38
AAG16483
ID AAG16483 standard; Protein; 237 AA.
XX
XX AAG16483;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 17146.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX

```

```

XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 14-JUN-1999; 99US-0138847.
XX 16-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX

```


PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 165;
Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
Db 119 peeeeeee 126
RESULT 36
AAR29716
ID AAR29716 standard; Protein; 232 AA.
XX
AC AAR29716;
XX
DT 30-APR-1993 (first entry)
XX
DE Guinea pig MBP-1.
XX
KW Major basic protein; anti-asthma.
XX
OS Cavia porcellus.
XX
FH Key Location/Qualifiers
FT Peptide 1..15
FT Peptide /note= "signal peptide"
FT Peptide 16..113
FT Protein /note= "prepeptide"
FT Protein 114..232
XX /note= "mature MBP-1"
XX JP04320686-A.
XX
PD 11-NOV-1992.
XX
PF 22-APR-1991; 91JP-0090704.
XX
PR 22-APR-1991; 91JP-0090704.
XX
PA (SAKA) OTSUKA PHARM CO LTD.
XX
XX WPI; 1992-426680/52.
DR N-PSDB; AAQ31988.
XX
PT A guinea pig MBP gene - contains DNA base sequence coding MBP
PT gene having 1 of 2 specified aminoacid sequences
XX
PS Disclosure; Page 15; 20pp; Japanese.
XX
CC The sequence shows the protein sequence of the guinea pig major basic
CC protein (MBP)-1. The sequence was obtd. by first isolating MBP-1
CC and MBP-2 from the abdominal transudate of male guinea pigs treated
CC with polymyxin and ascariis extract. MBP-1 and MBP-2 were used to
CC design probes for screening of a guinea pig eosinophil cDNA library.
CC The cDNA encoding the MBPs was used to transform suitable host
CC cells which may then be cultured to obtain large amts. of MBP.
CC The MBPs may be used for the confirmation of the effect of an anti-
CC asthma drug. See also AAR29717.
XX
SQ Sequence 232 AA;
Query Match 0.6%; Score 8; DB 13; Length 232;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
Db 72 peeeeeee 79
RESULT 37
AAW40075
ID AAW40075 standard; Protein; 233 AA.
XX

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 159;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
|
Db 129 peeeeee 136

RESULT 35

AAG39616
ID AAG39616 standard; Protein; 165 AA.

XX AC AAG39616;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 49044.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.


```
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 23-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161040.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 158;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
Db 90 peeeeeee 97

RESULT 34
AAG37860
ID RAG37860 standard; Protein; 159 AA.
XX
AC AAG37860;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46624.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
```


PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 0.6%; Score 8; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

DB 119 peeeeeee 126

RESULT 33

AAG31670
ID AAG31670 standard; Protein; 158 AA.

XX AC AAG31670;

XX DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38072.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.


```
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      0.6%; Score 8; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331.
Db 119 peeeeee 126

RESULT 32
AAG37861
ID AAG37861 standard; Protein; 149 AA.
XX
AC AAG37861;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46625.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 18-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
```


PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 25-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 81;

Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

|||||||

Db 51 peeeeee 58

RESULT 30

AAG39617

ID AAG39617 standard; Protein; 97 AA.

XX AC AAG39617;

XX XX

DT 18-OCT-2000 (first entry)

XX XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49045.

XX XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX XX

PN EP1033405-A2.

XX XX

PD 06-SEP-2000.

XX XX

PF 25-FEB-2000; 2000EP-0301439.

XX XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 81;

Best Local Similarity 100.0%; Pred.No.5.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

|||||||

Db 51 peeeeee 58

RESULT 29

AAG37862

ID AAG37862 standard; Protein; 81 AA.

XX AC AAG37862;

XX AC AAG37862;

DT 18-OCT-2000 (first entry)

XX DE

DE Arabidopsis thaliana protein fragment SEQ ID NO: 46626.

XX KW

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW

OS Arabidopsis thaliana.

XX OS

XX EP1033405-A2.

PN

XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PF 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134236.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136031.

PR 27-MAY-1999; 99US-0136382.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

99US-0140354.

99US-0140695.

99US-0140823.

99US-0140991.

99US-0141287.

99US-0141842.

99US-0142154.

99US-0142055.

99US-0142390.

99US-0142803.

99US-0142920.

99US-0142977.

99US-0143542.

99US-0143624.

99US-0144005.

99US-0144085.

99US-0144086.

99US-0144325.

99US-0144331.

99US-0144332.

99US-0144333.

99US-0144334.

99US-0144335.

99US-0144352.

99US-0144632.

99US-0144884.

99US-0144814.

99US-0145086.

99US-0145088.

99US-0145085.

99US-0145087.

99US-0145089.

99US-0145192.

99US-0145145.

99US-0145218.

99US-0145224.

99US-0145276.

99US-0145913.

99US-0145918.

99US-0145919.

99US-0145951.

99US-0146386.

99US-0146388.

99US-0146389.

99US-0147038.

99US-0147204.

99US-0147302.

99US-0147192.

99US-0147260.

99US-0147303.

99US-0147416.

99US-0147493.

99US-0147935.

99US-0148171.

99US-0148319.

99US-0148341.

99US-0148565.

99US-0148684.

99US-0149368.

99US-0149175.

99US-0149426.

99US-0149722.

99US-0149723.

99US-0149929.

99US-0149902.

99US-0149930.

99US-0150566.

99US-0150884.

99US-0151065.

99US-0151066.

99US-0151080.

99US-0151303.

99US-0151438.

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 8; DB 21; Length 74;
Best Local Similarity 100.0%; Pred. No. 4.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 42 eeeeeee 49

RESULT 28

AAG19544

ID AAG19544 standard; Protein; 81 AA.

XX AC AAG19544;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21385.
XX


```
Db      13 eeeeeer 20
RESULT  26
AAG60357 standard; Protein; 58 AA.
XX      AC      AAG60357;
XX      DT      18-OCT-2000 (first entry)
XX      DE      Arabidopsis thaliana protein fragment SEQ ID NO: 78170.
XX      KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX      OS      Arabidopsis thaliana.
XX      PN      EP1033405-A2.
XX      PD      06-SEP-2000.
XX      PF      25-FEB-2000; 2000EP-0301439.
XX      PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      05-MAY-1999; 99US-0132484.
PR      06-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      07-MAY-1999; 99US-0132487.
PR      11-MAY-1999; 99US-0132863.
PR      14-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      18-MAY-1999; 99US-0134370.
PR      19-MAY-1999; 99US-0134768.
PR      20-MAY-1999; 99US-0134941.
PR      21-MAY-1999; 99US-0135124.
PR      24-MAY-1999; 99US-0135353.
PR      25-MAY-1999; 99US-0135629.
PR      27-MAY-1999; 99US-0136021.
PR      28-MAY-1999; 99US-0136392.
PR      01-JUN-1999; 99US-0136782.
PR      03-JUN-1999; 99US-0137222.
PR      04-JUN-1999; 99US-0137528.
PR      07-JUN-1999; 99US-0137502.
PR      08-JUN-1999; 99US-0137724.
PR      10-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      14-JUN-1999; 99US-0138847.
PR      16-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      17-JUN-1999; 99US-0139453.
PR      18-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      21-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145216.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      03-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      04-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      05-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147303.
PR      06-AUG-1999; 99US-0147416.
PR      09-AUG-1999; 99US-0147493.
PR      09-AUG-1999; 99US-0147935.
PR      10-AUG-1999; 99US-0148171.
PR      11-AUG-1999; 99US-0148319.
PR      12-AUG-1999; 99US-0148341.
PR      13-AUG-1999; 99US-0148565.
PR      13-AUG-1999; 99US-0148684.
PR      16-AUG-1999; 99US-0149368.
```


PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.68; Score 8; DB 21; Length 58;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332

SQ Sequence 28 AA;

Query Match 0.6%; Score 8; DB 14; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
| | | | | | | |
Db 14 eeeeeer 21

RESULT 24

AAR95846
ID AAR95846 standard; Protein; 32 AA.XX
AC AAR95846;XX
DT 28-OCT-1996 (first entry)XX
DE Human prepro-TRH peptide having CRIF activity (aa 152-183).

XX TRH; thyrotropin release factor; Cushing's disease; anxiety;
KW corticotropin release inhibiting factor; CRIF; depression; obesity;
KW anorexia nervosa; withdrawal; hypocortisolism; colitis; autoimmune;
KW arthritis; premenstrual syndrome; inflammatory; obsessive compulsive;
KW disorder.

XX
OS Homo sapiens.XX
FH Key Location/Qualifiers

FT Modified-site 1

FT /label= OTHER

FT /note= "pyroglutamic acid"

FT Misc-difference 1..6

FT /note= "residues 1-6 may be opt. cleaved

FT i.e. absent"

XX
PN WO9608265-A1.XX
PD 21-MAR-1996.XX
PF 08-SEP-1995; 95WO-US11455.XX
PR 12-SEP-1994; 94US-0304383.XX
PA (UYPE-) UNIV PENNSYLVANIA.XX
PI Aird F, Redel E;XX
XX WPI; 1996-179720/18.DR
DR N-PSDB; AAT15289.

XX
PT New isolated corticotropin release inhibiting factor peptide(s) -
PT used to develop prods. for the diagnosis and treatment of CRIF
PT related disorders, e.g. stress responses or inflammation

XX
PS Claim 7; Figure 10; 66pp; English.

XX
CC AAR95846 is a portion (amino acids 152-183) of the human prepro-TRH
CC (thyrotropin releasing hormone) protein positioned between the
CC fourth and fifth TRH sequences. The peptide has CRIF (corticotropin
CC release inhibiting factor) activity and may be used for the
CC diagnosis and treatment of CRIF related disorders. Such disorders
CC include Cushing's disease, anxiety, anorexia nervosa, depression,
CC obesity, withdrawal from drug or alcoholic dependence, some cancers,
CC hypercortisolism, ACTH (adrenocorticotropin) deficiency; premenstrual
CC syndrome, inflammatory conditions e.g. colitis and autoimmune disease
CC e.g. arthritis. Peptides with CRIF activity may also be used to
CC enhance immune responses and to increase the efficacy of immunity to
CC vaccines.

XX
SQ Sequence 32 AA;

Query Match 0.6%; Score 8; DB 17; Length 32;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
| | | | | | | |
Db 18 eeeeeer 25

RESULT 25

AAG56078
ID AAG56078 standard; Protein; 58 AA.XX
AC AAG56078;XX
DT 18-OCT-2000 (first entry)XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 72021.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.XX
PN EP1033405-A2.XX
PD 06-SEP-2000.XX
PF 25-FEB-2000; 2000EP-0301439.XX
PR 25-FEB-1999; 99US-0121825.PR
PR 05-MAR-1999; 99US-0123180.PR
PR 09-MAR-1999; 99US-0123548.PR
PR 23-MAR-1999; 99US-0125788.PR
PR 25-MAR-1999; 99US-0126264.PR
PR 29-MAR-1999; 99US-0126785.PR
PR 01-APR-1999; 99US-0127462.PR
PR 06-APR-1999; 99US-0128234.PR
PR 08-APR-1999; 99US-0128714.PR
PR 16-APR-1999; 99US-0129845.PR
PR 19-APR-1999; 99US-0130077.PR
PR 21-APR-1999; 99US-0130449.PR
PR 23-APR-1999; 99US-0130510.PR
PR 28-APR-1999; 99US-0130891.PR
PR 30-APR-1999; 99US-0131449.PR
PR 30-APR-1999; 99US-0132048.PR
PR 04-MAY-1999; 99US-0132407.PR
PR 05-MAY-1999; 99US-0132484.PR
PR 06-MAY-1999; 99US-0132485.PR
PR 06-MAY-1999; 99US-0132486.PR
PR 07-MAY-1999; 99US-0132487.PR
PR 11-MAY-1999; 99US-0132863.PR
PR 14-MAY-1999; 99US-0134256.PR
PR 14-MAY-1999; 99US-0134218.PR
PR 14-MAY-1999; 99US-0134221.PR
PR 14-MAY-1999; 99US-0134370.PR
PR 18-MAY-1999; 99US-0134768.PR
PR 19-MAY-1999; 99US-0134941.PR
PR 20-MAY-1999; 99US-0135124.PR
PR 21-MAY-1999; 99US-0135353.PR
PR 24-MAY-1999; 99US-0135629.PR
PR 25-MAY-1999; 99US-0136021.PR
PR 27-MAY-1999; 99US-0136392.PR
PR 28-MAY-1999; 99US-0136782.PR
PR 01-JUN-1999; 99US-0137222.PR
PR 03-JUN-1999; 99US-0137528.PR
PR 04-JUN-1999; 99US-0137502.PR
PR 07-JUN-1999; 99US-0137724.PR
PR 08-JUN-1999; 99US-0138094.

KW post-traumatic stress disorder; therapy.

XX Homo sapiens.

XX US6039956-A.

XX 21-MAR-2000.

XX 07-JUN-1996; 96US-0660561.

XX 12-SEP-1994; 94US-0304383.

XX 08-SEP-1995; 95US-0523125.

XX (UYPE-) UNIV PENNSYLVANIA.

XX McGivern RF, Rittenhouse PA, Aird F, Redei E;

XX WPI; 2000-270120/23.

XX N-PSDB; AAA07229.

XX Treating behavioural symptoms including panic disorder, post-traumatic stress disorder and obsessive compulsive disorder in humans, involves administering corticotropin release inhibiting factor -

XX Claim 5; Column 35; 43pp; English.

XX This sequence is a human corticotropin release inhibitory factor (CRIF) fragment. The invention relates to a method for treating (I) behavioural symptoms in an anxiety disorder in a human, comprising administering CRIF which comprises 3 contiguous amino acids contained within the amino acid sequence positioned between the fourth and fifth thyrotropin releasing hormone (TRH) sequence on a prepro-TRH protein. (I) is useful for treating behavioural symptoms in an anxiety disorder which includes general anxiety disorder, panic disorder, obsessive compulsive disorder and post-traumatic stress disorder. Peripheral administration of CRIF is capable of crossing the blood-brain barrier.

XX Sequence 26 AA;

Query Match 0.6%; Score 8; DB 21; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 12 eeeeeer 19

RESULT 22

AAB61357

ID AAB61357 standard; peptide; 26 AA.

XX AAB61357;

XX 03-APR-2001 (first entry)

XX Human corticotropin release inhibiting factor peptide.

XX Depressive disorder; corticotropin release inhibiting factor; CRIF; bipolar; dysthymia; cyclothymia; premenstrual.

XX Homo sapiens.

XX WO200100222-A1.

XX 04-JAN-2001.

XX 09-JUN-2000; 2000WO-US16107.

XX 24-JUN-1999; 99US-0140962.

XX 04-AUG-1999; 99US-0366981.

XX

PA (NOUN) UNIV NORTHWESTERN.

XX Redei E;

XX WPI; 2001-091671/10.

XX Treating depressive disorder such as major and minor depression, bipolar disorders, dysthymia, cyclothymia, and premenstrual syndrome, involves administering a compound having corticotropin release inhibiting factor activity -

XX Claim 15; Page 28; 36pp; English.

XX The present invention relates to treating a depressive disorder in an animal, involving administering a compound with corticotropin release inhibiting factor (CRIF) activity. The invention is useful for treating depressive disorders such as major depression, minor depression, bipolar disorders, dysthymia, cyclothymia and premenstrual syndrome, in human.

XX Sequence 26 AA;

Query Match 0.6%; Score 8; DB 22; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

Db 12 eeeeeer 19

RESULT 23

AAR31234

ID AAR31234 standard; peptide; 28 AA.

XX AAR31234;

XX 19-MAY-1993 (first entry)

XX Prepro-thyrotropin releasing hormone related peptide.

XX Prepro TRH; antibodies; human; pituitary dysfunction; psychoses; thyroidal diseases; thyroid.

XX Synthetic.

XX JF04352797-A.

XX 07-DEC-1992.

XX 29-MAY-1991; 91JP-0228139.

XX 29-MAY-1991; 91JP-0228139.

XX (SAKA) OTSUKA PHARM CO LTD.

XX WPI; 1993-024099/03.

XX Human prepro-TRH related peptide(s) - used to prepare antibodies which react specifically with human prepro-TRH for diagnosis of e.g. pituitary dysfunction etc.

XX Claim 5; Page 16; 19pp; Japanese.

XX The sequence shown is that of a fragment of human prepro-thyrotropin releasing hormone (TRH) which may be prep. synthetically. The peptide can be used to raise antibodies which specifically react with human prepro-TRH. The resultant antibodies may be used for the diagnosis of human TRH related diseases, e.g. pituitary dysfunction, thyroidal diseases and psychoses.

XX See also AAR31229-36.

XX


```

SQ Sequence 1886 AA;
Query Match 0.6%; Score 9; DB 19; Length 1886;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKOLEV 565
Db 1476 ekirkqlev 1484
|||||

RESULT 19
AAR31233
ID AAR31233 standard; peptide; 26 AA.
XX
AC AAR31233;
XX
DT 19-MAY-1993 (first entry)
XX
DE Prepro-thyrotropin releasing hormone related peptide.
XX
KW Prepro TRH; antibodies; human; pituitary dysfunction; psychoses;
KW thyroidal diseases; thyroid.
XX
OS Synthetic.
XX
PN JP04352797-A.
XX
PD 07-DEC-1992.
XX
PF 29-MAY-1991; 91JP-0228139.
XX
PR 29-MAY-1991; 91JP-0228139.
XX
PA (SAKA ) OTSUKA PHARM CO LTD.
XX
DR WPI; 1993-024099/03.
XX
PT Human prepro-TRH related peptide(s) - used to prepare antibodies
PT which react specifically with human prepro-TRH for diagnosis of
PT e.g. pituitary dysfunction etc.
XX
PS Claim 4; Page 16; 19pp; Japanese.
XX
CC The sequence shown is that of a fragment of human prepro-thyrotropin
CC releasing hormone (TRH) which may be prep. synthetically. The
CC peptide can be used to raise antibodies which specifically react with
CC human prepro-TRH. The resultant antibodies may be used for the
CC diagnosis of human TRH related diseases, e.g. pituitary dysfunction,
CC thyroidal diseases and psychoses.
CC See also AAR31229-36.
XX
SQ Sequence 26 AA;

Query Match 0.6%; Score 8; DB 14; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db 12 eeeeeer 19
|||||

RESULT 20
AAW73133
ID AAW73133 standard; peptide; 26 AA.
XX
AC AAW73133;
XX
DT 15-JAN-1999 (first entry)
XX
DE Human CRIF protein sequence fragment.
XX
KW Corticotropin release inhibitory factor; CRIF; behavioural symptom;
KW anxiety disorder; thyrotropin releasing hormone; TRH sequence; human;
KW general anxiety disorder; panic disorder; obsessive compulsive disorder;

```

```

DE XX Corticotropin release inhibiting factor peptide.
KW CRIF; corticotropin release inhibiting factor; Cushing's disease; cancer;
KW hypothalamic corticotropin-releasing factor; anorexia nervosa; obesity;
KW pituitary adrenocorticotropin; adrenocortical glucocorticoid; depression;
KW thyrotropin-releasing hormone; CRIF disorder; anxiety; viral infection;
KW therapy.
XX
OS Homo sapiens.
XX
PN US5830866-A.
XX
PD 03-NOV-1998.
XX
PF 08-SEP-1995; 95US-0523125.
XX
PR 08-SEP-1995; 95US-0523125.
PR 12-SEP-1994; 94US-0304383.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI Aird F, Redei E;
XX
DR WPI; 1998-609292/51.
XX
PT Reduction of e.g. pituitary adrenocorticotropin - comprises
PT administration of fragment of prepro thyrotropin-releasing hormone
PT protein
XX
PS Claim 7; Column 26; 33pp; English.
XX
CC This sequence represents a corticotropin release inhibiting factor (CRIF)
CC peptide. This CRIF peptide can be used in the method of the invention for
CC reducing the level of hypothalamic corticotropin-releasing factor,
CC pituitary adrenocorticotropin and/or adrenocortical glucocorticoid
CC comprises administration of a CRIF suspended in a carrier. CRIF is a
CC recombinant or synthetic peptide comprising at least three contiguous
CC amino acids from the amino acid sequence positioned between the fourth
CC and fifth thyrotropin-releasing hormone (TRH) sequences of a prepro-TRH
CC protein. The method is used for the treatment of CRIF disorders
CC characterised by elevated levels of hypothalamic corticotropin-releasing
CC factor, pituitary adrenocorticotropin and/or adrenocortical
CC glucocorticoid. Such disorders include Cushing's disease,
CC anxiety, anorexia nervosa, depression, obesity, alcohol and drug
CC withdrawal symptoms, disorders exacerbated by stress-induced
CC immunosuppression (e.g. viral infections), and cancers whose severity is
CC reduced by an enhanced immune response.
XX
SQ Sequence 26 AA;

Query Match 0.6%; Score 8; DB 19; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
Db 12 eeeeeer 19
|||||

RESULT 21
AAAY81900
ID AAAY81900 standard; Peptide; 26 AA.
XX
AC AAAY81900;
XX
DT 22-JUN-2000 (first entry)
XX
DE Human CRIF protein sequence fragment.
XX
KW Corticotropin release inhibitory factor; CRIF; behavioural symptom;
KW anxiety disorder; thyrotropin releasing hormone; TRH sequence; human;
KW general anxiety disorder; panic disorder; obsessive compulsive disorder;

```


RESULT 17
 ID AAB01832 standard; Protein; 1011 AA.
 XX AC AAB01832;
 XX DT 11-SEP-2000 (first entry)
 XX DE Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.
 XX KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis.
 XX OS Haemophilus influenzae strain K21.
 XX PN WO200020609-A2.
 XX PD 13-APR-2000.
 XX PF 07-OCT-1999; 99WO-CA00938.
 XX PR 07-OCT-1998; 98US-0167568.
 XX PR 08-DEC-1998; 98US-0206942.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Loosmore SM, Yang Y, Klein MH;
 XX DR WPI: 2000-303789/26.
 XX DR N-PSDB; AAA52181.
 XX KW Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX
 PS Claim 12; Fig 21A-O; 307pp; English.
 XX
 CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
 CC strains Joyce, K1, K21, LDCP2, PWH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMWA protein from a non-typeable strain of
 CC H. influenzae.
 XX SQ Sequence 1011 AA;

Query Match 0.6%; Score 9; DB 21; Length 1011;

Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 EFKLKSNDN 286
 |||||
 Db 269 efklksndn 277

RESULT 18
 AAW54241
 ID AAW54241 standard; Protein; 1886 AA.
 XX AC AAW54241;
 XX DT 17-AUG-1998 (first entry)
 XX DE Rattus norvegicus mutant alpha-myosin heavy chain.

XX Actin binding domain; alpha myosin heavy chain; bridging; mouse;
 KW transgenic mammal; congestive heart failure; study; treatment; diet;
 KW exercise; effects; identification; hypertrophic cardio-myopathy;
 KW dilated or hypertrophic cardiomyopathy; acute aortic regurgitation;
 KW tricuspid stenosis; constrictive pericarditis; hypertension;
 KW acute infective endocarditis; ischaemic heart disease;
 KW primary myocardial disease; valvular disease; pericardial disease;
 KW hyperthyroidism; anaemia; arteriovenous fistula; beri-beri;
 KW Paget's disease; transgene.

OS Rattus norvegicus.
 OS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Region 403
 XX FT /note= "Arg403Gln mutation"
 XX FT Region 468..475
 XX FT /note= "nonmyosin actin binding domain bridge"

XX WO9813476-A1.

XX 02-APR-1998.

XX 26-SEP-1997; 97WO-US17296.

XX 26-SEP-1996; 96US-0026742.

XX (LEIN/) LEINWAND LA.

XX Vikstrom KL;

XX WPI: 1998-230690/20.

XX N-PSDB; AAV21518.

XX Transgenic mouse models for congestive heart failure and
 PT hypertrophic cardio-myopathy - used to study molecular and cellular
 PT events, identify potential therapeutic agents, assess effects of
 PT diet etc.

XX Disclosure; Pages 53-58; 75pp; English.

XX The sequence is that of a mutant rat alpha-myosin heavy chain
 CC which was used in the development of transgenic mammals,
 CC specifically mice. They can be used as a model for studying congestive
 CC heart failure (CHF) or hypertrophic cardiomyopathy. Such animals
 CC are used to study molecular and cellular events associated with
 CC CHF; to identify compounds for treating CHF, and in evaluating
 CC effects of diet and exercise on CHF. Conditions associated with CHF
 CC that can be evaluated this way are dilated or hypertrophic
 CC cardiomyopathy; acute aortic regurgitation; tricuspid stenosis;
 CC constrictive pericarditis; acute infective endocarditis; ischaemic
 CC heart disease; hypertension; primary myocardial disease; valvular
 CC or pericardial disease; hyperthyroidism; anaemia; arteriovenous
 CC fistula; beri-beri and Paget's disease.

SQ Sequence 439 AA;

Query Match 0.6%; Score 9; DB 21; Length 439;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1331
 |||||
 Db 53 apeeeee 61

RESULT 15
 AAB66298
 ID AAB66298 standard; Protein; 439 AA.
 XX
 AC AAB66298;
 XX
 DT 05-APR-2001. (first entry)
 XX
 DE Human TRF1 SEQ ID NO: 151.
 XX
 KW Human; tankyrase2; TANK2; TRF1; telomere; cancer; neoplasm; aging;
 KW inflammatory disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200100849-A1.
 XX
 PD 04-JAN-2001.
 XX
 PF 28-JUN-2000; 2000WO-US17827.
 XX
 PR 29-JUN-1999; 99US-0141582.
 XX
 PA (ICOS-) ICOS CORP.
 XX
 PI Christenson E, Demaggio AJ, Goldman PS, McElligott DL;
 XX
 DR WPI; 2001-102896/11.
 DR N-PSDB; AAF63966.
 XX
 PT New tankyrase2 polypeptides, useful for treating conditions mediated by
 PT poly(adenosine diphosphate-ribose) polymerase activity e.g. cancers,
 PT inflammatory and autoimmune disorders -
 XX
 PS Example 6; Page 223-225; 242pp; English.
 XX
 CC The present invention provides the protein and coding sequence for the
 CC human tankyrase2 protein. This is found in two different versions,
 CC designated TANK2-LONG and TANK2-SHORT. Tankyrase2 has
 CC polyADP-ribosylation activity and is involved in the modification of
 CC TRF1, which is a telomere-specific binding protein. The regulation of
 CC telomere length, in which TRF1 has a role, is linked to ageing and
 CC cancer. The sequences are useful in the treatment of cancers and
 CC inflammatory disorders.
 XX
 SQ Sequence 439 AA;

Query Match 0.6%; Score 9; DB 22; Length 439;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1331
 |||||
 Db 53 apeeeee 61

RESULT 16
 AAB01833
 ID AAB01833 standard; Protein; 1005 AA.
 XX

AC AAB01833;
 XX
 DT 11-SEP-2000 (first entry)
 XX
 DE Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.
 XX
 KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis.
 XX
 OS Haemophilus influenzae strain K21.
 XX
 PN WO200020609-A2.
 XX
 PD 13-APR-2000.
 XX
 PF 07-OCT-1999; 99WO-CA00938.
 XX
 PR 07-OCT-1998; 98US-0167568.
 XX
 PR 08-DEC-1998; 98US-0206942.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 XX
 XX Loosmore SM, Yang Y, Klein MH;
 XX
 XX WPI: 2000-303789/26.
 XX
 DR N-PSDB; AAS2182.
 DR
 XX
 PT Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX
 XX Claim 8; Fig 21A-O; 307pp; English.
 XX
 CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMW. The invention also discloses hmwA genes (AAS2175-A52198)
 CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
 CC strains Jovc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents a mature HMW protein from a non-typeable
 CC strain of H. influenzae.
 XX
 SQ Sequence 1005 AA;

Query Match 0.6%; Score 9; DB 21; Length 1005;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 EFKLKSDND 286
 |||||
 Db 263 efklksndn 271

PS Disclosure; Page 112-114; 163pp; English.

XX The telomere repeat binding factor (TRF) is used in the production of an
 CC altered vertebrate telomere repeat binding protein (A-TRF) which has a
 CC TRF dimerisation domain, and forms a hetero-dimer with TRF, preventing it
 CC from binding to the specified repeat sequence. A-TRF, optionally
 CC expressed by gene therapy, is used to inhibit shortening of telomeres
 CC associated with ageing (for cosmetic purposes) and disease, e.g. ataxia
 CC telangeiectasia, Down's syndrome, atrophy of the skin, age-related macular
 CC degeneration, atherosclerosis, tumours and viral (including human immune
 CC deficiency virus) infection. Cells expressing A-TRF also have an
 CC increased life span in vitro, e.g. for expression of recombinant proteins
 CC or where intended for subsequent transplant or for testing, eliminating
 CC the need for transformation.

SQ Sequence 439 AA;

Query Match 0.6%; Score 9; DB 19; Length 439;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331
 |
 Db 53 apeeeeeee 61

RESULT 13

ID AAY02182
 AC AAY02182 standard; Protein; 439 AA.

XX AAY02182;

XX 07-JUL-1999 (first entry)

XX A human telomere repeat binding factor (TRF).

XX Telomere repeat binding factor; TRF; TRF2; telomere termini;
 KW cell proliferation; antagonist; agonist; telomere length reduction;
 KW aging; skin atrophy; age-related macular degeneration; atherosclerosis;
 KW abnormal telomere length; cancer cell; telomere-associated disorder;
 KW mammalian artificial chromosome; gene therapy; viral infection.

XX Homo sapiens.

XX WO9915662-A1.

XX 01-APR-1999.

XX 25-SEP-1998; 98WO-US20175.

XX 04-FEB-1998; 98US-0018635.

XX 26-SEP-1997; 97US-0938052.

XX (UYRQ) UNIV ROCKEFELLER.

XX Broccoli D, De Lange T, Smogorzenska A;

XX WPI; 1999-263481/22.

XX N-PSDB; AAX35386.

XX Nucleic acid encoding telomere repeat binding factor for treatment
 of, e.g. cancer

XX Disclosure; Fig 3; 160pp; English.

XX The present sequence represents human telomere repeat binding
 CC factor (TRF). The specification also describes vertebrate
 CC TRF2 polynucleotides and polypeptides. When the basic N-terminal
 CC domain of the TRF2 protein is removed, it binds detectably to the
 CC telomere repeat sequence (TTAGGG)₁₂. TRF maintains the correct
 CC structure of telomere termini and protects against end-to-end fusion.
 CC It is required for cell proliferation. TRFs (or their antagonists and

CC agonists) can be used to limit reduction in telomere length associated
 CC with aging (e.g. atrophy of the skin, age-related macular degeneration
 CC and atherosclerosis) or abnormal telomere lengths in cancer cells.
 CC They can also be used to screen for specific modulators (potentially
 CC useful for treating aging and cancer), including those that are specific
 CC for one TRF over another; to diagnose telomere-associated disorders; as
 CC targeting agents for TTAGG repeats and in construction of mammalian
 CC artificial chromosomes (for gene therapy or basic research). Antibodies
 CC against TRF can be used for diagnosis and therapy, e.g. to differentiate
 CC between different TRFs to screen expression libraries for TRF-expressing
 CC genes or to detect (pre)cancers or viral infections.

XX Sequence 439 AA;

Query Match 0.6%; Score 9; DB 20; Length 439;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331
 |
 Db 53 apeeeeeee 61

RESULT 14

ID AAB03942

XX AAB03942 standard; Protein; 439 AA.

XX AAB03942;

XX 26-FEB-2001 (first entry)

XX Human telomeric repeat binding factor 1.

XX Human telomeric repeat binding factor 1; hTRBF-1; antisense;
 KW disease; modulation; expression; prophylaxis; infection;
 KW inflammation; anti-inflammatory; tumour; diagnostic; human.

XX Homo sapiens.

XX US6130088-A.

XX 10-OCT-2000.

XX 21-JUL-1999; 99US-0358384.

XX 21-JUL-1999; 99US-0358384.

XX (ISIS-) ISIS PHARM INC.

XX Monia BP, Cowser LM;

XX WPI; 2000-664192/64.

XX N-PSDB; AAA54235.

XX New antisense compounds that hybridizes with and inhibits the
 PT expression of human telomeric repeat binding factor 1 (TRBF-1), useful
 PT for treating conditions or diseases associated with TRBF-1 expression
 XX Example 1; Columns 47-50; 34pp; English.

XX Antisense compounds directed against the start codon, 3' untranslated
 CC region, nucleotides 78-374, 560-681, 965-1334 of the coding region or
 CC the stop codon of the human telomeric repeat binding factor 1
 CC (hTRBF-1) can be used to inhibit the expression of hTRBF-1. The
 CC antisense compounds are used for treating a patient suspected of
 CC having or being prone to a disease or condition associated with
 CC expression of TRBF-1 by modulating its expression. They may also be
 CC used prophylactically to prevent or delay infection, inflammation or
 CC tumour formation, or as research reagents and diagnostics, e.g. to
 CC distinguish between functions of various members of a biological
 CC pathway.

DT 14-DEC-1998 (first entry)
 XX
 DE Altered telomere repeat binding factor protein.
 XX
 KW human; telomere repeat binding factor; A-TRF; dimerisation domain;
 XX telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO9836066-A1.
 XX
 PD 20-AUG-1998.
 XX
 XX 13-FEB-1998; 98WO-US02765.
 XX
 PR 04-FEB-1998; 98US-0018628.
 PR 13-FEB-1997; 97US-0800264.
 XX
 PA (UYRQ) UNIV ROCKEFELLER.
 XX
 XX Bianchi A, De Lange T, Van Steensel B;
 XX
 XX WPI: 1998-480769/41.
 DR N-PSDB; AAV59280.
 DR
 XX
 PT Nucleic acid encoding altered telomere repeat binding protein and
 PT related vectors - transformants, hetero-dimers and antibodies, used
 PT to inhibit shortening of telomerases caused by ageing or disease,
 PT also used to extend life of cells in culture
 XX
 PS Claim 13; Page 115-117; 163pp; English.
 XX
 CC The altered vertebrate telomere repeat binding protein (A-TRF) has a
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to
 CC inhibit shortening of telomeres associated with ageing (for cosmetic
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,
 CC tumours and viral (including human immune deficiency virus) infection.
 CC Cells expressing A-TRF also have an increased life span in vitro, e.g.
 CC for expression of recombinant proteins or where intended for subsequent
 CC transplant or for testing, eliminating the need for transformation.
 XX
 SQ Sequence 437 AA;
 XX
 Query Match 0.6%; Score 9; DB 19; Length 437;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1323 APEEEEEEE 1331
 Db 53 apeeeeeee 61
 RESULT 11
 AAW28555
 ID AAW28555 standard; Protein; 439 AA.
 XX
 AC AAW28555;
 XX
 XX 09-FEB-1998 (first entry)
 DT
 DE Telomeric repeat binding factor.
 XX
 KW Telomeric repeat binding factor; TRF; ageing; cancer.
 XX
 OS Homo sapiens.
 XX
 XX WO9708314-A2.
 PN
 XX

PD 06-MAR-1997.
 XX
 PF 23-AUG-1996; 96WO-US13652.
 XX
 PR 25-AUG-1995; 95US-0519103.
 XX
 PA (UYRQ) UNIV ROCKEFELLER.
 XX
 XX De LANGE T;
 XX
 XX WPI: 1997-179271/16.
 DR N-PSDB; AAT87489.
 DR
 XX Telomeric repeat binding factor of mol. wt. 67 kD - useful for
 PT diagnosis and treatment of cellular ageing and cancer
 PT
 XX Claim 4; Fig 3; 62pp; English.
 XX
 CC TRF is the first telomeric protein isolated from a multicellular
 CC organism. Human telomeres shorten during normal cell division
 CC and this shortening may eventually limit cell proliferation
 CC and lead to ageing. Addition of TRF, its analogues, agonists
 CC or mimics will be useful for restoring telomere function to treat
 CC cellular ageing.
 XX
 SQ Sequence 439 AA;
 XX
 Query Match 0.6%; Score 9; DB 18; Length 439;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1323 APEEEEEEE 1331
 Db 53 apeeeeeee 61
 RESULT 12
 AAW77348
 ID AAW77348 standard; Protein; 439 AA.
 XX
 AC AAW77348;
 XX
 XX 14-DEC-1998 (first entry)
 DT
 DE Human telomere repeat binding factor protein.
 XX
 KW Human; telomere repeat binding factor; A-TRF; dimerisation domain;
 KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.
 XX
 OS Homo sapiens.
 XX
 PN WO9836066-A1.
 XX
 PD 20-AUG-1998.
 XX
 XX 13-FEB-1998; 98WO-US02765.
 PF
 PR 04-FEB-1998; 98US-0018628.
 PR 13-FEB-1997; 97US-0800264.
 XX
 XX (UYRQ) UNIV ROCKEFELLER.
 PA
 XX Bianchi A, De Lange T, Van Steensel B;
 PI
 XX WPI: 1998-480769/41.
 DR N-PSDB; AAV59278.
 DR
 XX Nucleic acid encoding altered telomere repeat binding protein and
 PT related vectors - transformants, hetero-dimers and antibodies, used
 PT to inhibit shortening of telomerases caused by ageing or disease,
 PT also used to extend life of cells in culture
 PT
 XX

CC cancer. The sequences are useful in the treatment of cancers and
 CC inflammatory disorders.

SQ Sequence 67 AA;

Query Match 0.6%; Score 9; DB 22; Length 67;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

Db 53 apeeeeeee 61
 |||||

RESULT 8
 AAG04041
 ID AAG04041 standard; Protein; 114 AA.

AC AAG04041;

DT 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 8122.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

PI WPI; 2000-500381/45.

DR N-PSDB; AAC04047.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 8122; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX SQ Sequence 114 AA;

Query Match 0.6%; Score 9; DB 21; Length 114;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

Db 53 apeeeeeee 61
 |||||

RESULT 9

AAW77351

ID AAW77351 standard; Protein; 325 AA.

AC AAW77351;

DT 14-DEC-1998 (first entry)

XX Altered telomere repeat binding factor protein deletion mutant.

XX Human; telomere repeat binding factor; A-TRF; dimerisation domain;
 KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

XX Homo sapiens.

XX Synthetic.

XX WO9836066-A1.

XX 20-AUG-1998.

PF 13-FEB-1998; 98WO-US02765.

XX 04-FEB-1998; 98US-0018628.

PR 13-FEB-1997; 97US-0800264.

XX (UYRQ) UNIV ROCKEFELLER.

XX Bianchi A, De Lange T, Van Steensel B;

DR WPI; 1998-480769/41.

XX N-PSDB; AAV59281.

XX Nucleic acid encoding altered telomere repeat binding protein and
 PT related vectors - transformants, hetero-dimers and antibodies, used
 PT to inhibit shortening of telomerases caused by ageing or disease,
 PT also used to extend life of cells in culture

XX Claim 15; Page 117-118; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to
 CC inhibit shortening of telomerases associated with ageing (for cosmetic
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,
 CC tumours and viral (including human immune deficiency virus) infection.
 CC Cells expressing A-TRF also have an increased life span in vitro, e.g.
 CC for expression of recombinant proteins or where intended for subsequent
 CC transplant or for testing, eliminating the need for transformation.

XX SQ Sequence 325 AA;

Query Match 0.6%; Score 9; DB 19; Length 325;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

Db 53 apeeeeeee 61
 |||||

RESULT 10

AAW77350

ID AAW77350 standard; Protein; 437 AA.

AC AAW77350;

XX

XX PS Claim 11; Page 1337-1339; 1425pp; English.

XX CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer

XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

XX CC associated proteins and polynucleotide sequences, their agonists, and

XX CC antagonists may have neuroprotective; cytostatic; cardioactive;

XX CC immunomodulatory; muscular active general; vulnary; gastrointestinal

XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial

XX CC activity. The invention also includes antibodies specific for the

XX CC protein or polynucleotide sequences. The lung cancer associated

XX CC polynucleotide sequences may be used for detection of lung cancer.

XX CC chromosome identification, as chromosome markers, and for numerous other

XX CC diagnostic or research purposes. The proteins may be used to treat

XX CC disorders such as neural, immune, muscular, reproductive,

XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

XX CC disorders. The proteins may also be used in the treatment of wounds and

XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

XX CC peptide AAB58549 are used in the course of the invention for the

XX CC identification and characterisation of the polynucleotide and protein

XX CC sequences.

XX SQ Sequence 363 AA;

Query Match 1.0%; Score 14; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 9.1e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 EKLISLLPEYVVPY 994

Db |||||||

17 eklisllpeyvvpv 30

RESULT 6

AAAY02188

ID AAY02188 standard; Protein; 67 AA.

XX AC AAY02188;

XX DT 07-JUL-1999 (first entry)

XX DE Polypeptide sequence from SEQ ID 33.

XX KW Telomere repeat binding factor; TRF; TRF2; telomere termini;

XX KW cell proliferation; antagonist; agonist; telomere length reduction;

XX KW aging; skin atrophy; age-related macular degeneration; atherosclerosis;

XX KW abnormal telomere length; cancer cell; telomere-associated disorder;

XX KW mammalian artificial chromosome; gene therapy; viral infection.

XX OS Homo sapiens.

XX PN WO9915662-A1.

XX PD 01-APR-1999.

XX PF 25-SEP-1998; 98WO-US20175.

XX PR 04-FEB-1998; 98US-0018635.

XX PR 26-SEP-1997; 97US-0938052.

XX PA (UYRQ) UNIV ROCKEFELLER.

XX PI Broccoli D, De Lange T, Smogorzenska A;

XX DR WPI; 1999-263481/22.

XX DR N-PSDB; AAX35393.

XX KW Nucleic acid encoding telomere repeat binding factor for treatment

XX PT of, e.g. cancer

XX PS Disclosure; Page 133; 160pp; English.

XX

CC The specification describes vertebrate telomere repeat binding

CC factor (TRF) and TRF2 polynucleotides and polypeptides. When the basic

CC N-terminal domain of the TRF protein is removed, it binds detectably to

CC the telomere repeat sequence (TTAGGG)₁₂. TRF maintains the correct

CC structure of telomere termini and protects against end-to-end fusion.

CC It is required for cell proliferation. TRFs (or their antagonists and

CC agonists) can be used to limit reduction in telomere length associated

CC with aging (e.g. atrophy of the skin, age-related macular degeneration

CC and atherosclerosis) or abnormal telomere lengths in cancer cells.

CC They can also be used to screen for specific modulators (potentially

CC useful for treating aging and cancer), including those that are specific

CC for one TRF over another; to diagnose telomere-associated disorders; as

CC targeting agents for TTAGGG repeats and in construction of mammalian

CC artificial chromosomes (for gene therapy or basic research). Antibodies

CC against TRF can be used for diagnosis and therapy, e.g. to differentiate

CC between different TRFs to screen expression libraries for TRF-expressing

CC genes or to detect (pre)cancers or viral infections. The present

CC sequence is used in the course of the invention.

XX SQ Sequence 67 AA;

Query Match 0.6%; Score 9; DB 20; Length 67;

Best Local Similarity 100.0%; Pred. No. 0.37;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEE 1331

Db |||||||

53 apeeeeeee 61

RESULT 7

AAB66299

ID AAB66299 standard; Protein; 67 AA.

XX AC AAB66299;

XX DT 05-APR-2001 (first entry)

XX DE Human TRF1 TANK1 binding domain protein sequence SEQ ID NO: 155.

XX KW Human; tankyrase2; TANK2; TRF1; telomere; cancer; neoplasm; aging;

XX KW inflammatory disorder.

XX OS Homo sapiens.

XX PN WO200100849-A1.

XX PD 04-JAN-2001.

XX PF 28-JUN-2000; 2000WO-US17827.

XX PR 29-JUN-1999; 99US-0141582.

XX PA (ICOS-) ICOS CORP.

XX PI Christenson E, Demaggio AJ, Goldman PS, McElligott DL;

XX DR WPI; 2001-102896/11.

XX DR N-PSDB; AAF63969.

XX PT New tankyrase2 polypeptides, useful for treating conditions mediated by

XX PT poly(adenosine diphosphate-ribose) polymerase activity e.g. cancers,

XX PT inflammatory and autoimmune disorders -

XX PS Example 6; Page 225-226; 242pp; English.

XX CC The present invention provides the protein and coding sequence for the

XX CC human tankyrase2 protein. This is found in two different versions,

XX CC designated TANK2-LONG and TANK2-SHORT. Tankyrase2 has

XX CC polyADP-ribosylation activity and is involved in the modification of

XX CC TRF1, which is a telomere-specific binding protein. The regulation of

XX CC telomere length, in which TRF1 has a role, is linked to ageing and

FT Modified-site 492 "Potential glycosylation site"

FT Modified-site 494 "Potential glycosylation site"

FT Modified-site 507 "Potential phosphorylation site"

FT Modified-site 507 "Potential phosphorylation site"

FT Modified-site 557 "Potential phosphorylation site"

FT Modified-site 557 "Potential phosphorylation site"

FT Modified-site 561 "Potential phosphorylation site"

FT Modified-site 561 "Potential phosphorylation site"

FT Modified-site 582 "Potential phosphorylation site"

FT Modified-site 582 "Potential phosphorylation site"

FT Modified-site 593 "Potential phosphorylation site"

FT Modified-site 593 "Potential phosphorylation site"

FT Modified-site 594 "Potential phosphorylation site"

FT Modified-site 594 "Potential phosphorylation site"

FT Modified-site 614 "Potential phosphorylation site"

FT Modified-site 614 "Potential phosphorylation site"

FT Modified-site 626 "Potential phosphorylation site"

FT Modified-site 626 "Potential phosphorylation site"

FT Modified-site 649 "Potential phosphorylation site"

FT Modified-site 649 "Potential phosphorylation site"

FT Modified-site 668 "Potential phosphorylation site"

FT Modified-site 668 "Potential phosphorylation site"

FT Modified-site 677 "Potential phosphorylation site"

FT Modified-site 677 "Potential phosphorylation site"

FT Modified-site 697 "Potential phosphorylation site"

FT Modified-site 697 "Potential phosphorylation site"

FT Modified-site 712 "Potential phosphorylation site"

FT Modified-site 712 "Potential phosphorylation site"

FT Modified-site 747 "Potential phosphorylation site"

FT Modified-site 747 "Potential phosphorylation site"

FT Modified-site 750 "Potential phosphorylation site"

FT Modified-site 750 "Potential phosphorylation site"

FT Modified-site 757 "Potential phosphorylation site"

FT Modified-site 757 "Potential phosphorylation site"

FT Modified-site 798 "Potential phosphorylation site"

FT Modified-site 798 "Potential phosphorylation site"

FT Modified-site 800 "Potential phosphorylation site"

FT Modified-site 800 "Potential phosphorylation site"

XX WO200034477-A2.

XX 15-JUN-2000.

XX 10-DEC-1999; 99WO-US30408.

XX 11-DEC-1998; 98US-0210083.

PR 11-DEC-1998; 98US-9123456.

PR 09-FEB-1999; 99US-0119365.

PR 16-MAR-1999; 99US-0124687.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;

PI Lu DAM, Azimzai Y;

XX WPI; 2000-423423/36.

DR N-PSDB; AAA47423.

XX New human neuron-associated proteins and polynucleotides encoding them,

PT useful for diagnosis, treatment and prevention of cell proliferative

PT disorders including cancer, neuronal and neurological disorders

XX Claim 1; Page 103-105; 145pp; English.

XX Human neuron-associated proteins (NEUAP) can be used for for

CC treating or preventing a disorder associated with decreased

CC expression or activity of NEUAP. Antagonists of NEUAP are useful for

CC treating or preventing disorder associated with increased expression

CC or activity of NEUAP. NEUAP or their fragments or derivatives are

CC useful for treating neurological disorder such as epilepsy, ischemic

CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's

CC disease, Pick's disease, Huntington's disease, dementia and

CC Parkinson's disease. NEUAPs are also useful for treating other

CC demyelinating diseases, bacterial and viral meningitis, prion

CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and

CC metabolic diseases of the nervous system, neurofibromatosis, other

CC developmental disorders of the central nervous system, cerebral

CC palsy, neuroskeletal disorders, autonomic nervous system disorders,

CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and

CC other neuromuscular disorders, peripheral nervous system disorders,

CC inherited, metabolic, endocrine, and toxic myopathies, mental

CC disorders including mood, anxiety and schizophrenic disorders, a cell

CC proliferative disorder such as actinic keratosis, arteriosclerosis,

CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective

CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,

CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory

CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's

CC disease, adult respiratory distress syndrome, allergies, ankylosing

CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,

CC complications of cancer, hemodialysis, and extracorporeal circulation,

CC viral, bacterial, fungal parasitic, protozoal, and helminthic

CC infections, and trauma. This protein was given the incyte ID no.

CC 2888437CDL.

XX SQ Sequence 854 AA;

Query Match 1.4%; Score 19; DB 21; Length 854;

Best Local Similarity 100.0%; Pred. No. 1e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 ALCADKDPVKERRAHARQCL 962

Db 471 alcakdpvkerraharqcl 489

|||||

RESULT 5

AAB58462

ID AAB58462 standard; Protein; 363 AA.

XX AC AAB58462;

XX 14-MAR-2001 (first entry)

XX Lung cancer associated polypeptide sequence SEQ ID 800.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;

XX cardioactive; immunomodulatory; muscular active; vulnerary;

XX gastrointestinal; nephrotropic; antiinfective; gynecological;

XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.

XX OS

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI; 2000-587514/55.

XX N-PSDB; AAF18338.

XX Lung cancer associated gene sequences, referred to as lung cancer

PT antigens, useful for treatment, prevention, and diagnosis of disorders

PT such as lung cancer -

Db 1 mknhsksgtstlrltllilhsdgdltteqgkiskpmsrlrlaagsaivklagepcyhei 60
QY 899 ITLEQYOLCALAINDECYQVQVFAQLKRG 929
Db 61 itleqyqlcalaindecyqvrqvfaklkg 91
RESULT 3
AAV65247
ID AAY65247 standard; Protein; 101 AA.
AC AAY65247;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST related polypeptide SEQ ID NO:1408.
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.
XX
OS Homo sapiens.
PN W09953051-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
DR WPI; 2000-038446/03.
DR N-PSDB; AA242861.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 3; Page 776; 837pp; English.
XX
CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AA242265 to
CC AA243052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 101 AA;

Query Match 5.2%; Score 72; DB 21; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-67;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSKTRTNDGKITYPPGVKEISDKISKEMVRRRLKNVVKTFMDMDQDSSEKELYNL 60
Db 1 mahsktrtdngkityppgvkeiskiskemvrrlknvktfmdmdqdsseekelylnla 60
QY 61 LHLASDFFLKHP 72
Db 61 lhlasdfflkhp 72
RESULT 4
AAB01382
ID AAB01382 standard; Protein; 854 AA.
XX
AC AAB01382;
XX
DT 20-OCT-2000 (first entry)
XX
DE Neuron-associated protein.
XX
KW Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW Alzheimer's disease; Pick's disease; Huntington's disease;
KW dementia; Parkinson's disease; demyelinating disease; meningitis;
KW prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW peripheral nervous system; PNS; myopathy; schizophrenia;
KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW AIDS; Addison's disease; adult respiratory distress syndrome;
KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW Werner syndrome; trauma; human.
XX
OS Homo sapiens.
XX
PH Key
FT Modified-site 48
FT Modified-site /note= "Potential glycosylation site"
FT Modified-site 88
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 99
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 153
FT Modified-site /note= "Potential glycosylation site"
FT Modified-site 159
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 175
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 249
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 265
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 330
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 340
FT Modified-site /note= "Potential phosphorylation site"
FT Region 358..409
FT Modified-site /label= "Postsynaptic protein signature"
FT Region 369
FT Modified-site /note= "Potential glycosylation site"
FT Region 373..422
FT Modified-site /label= "Neuraxin signature"
FT Modified-site 375
FT Modified-site /note= "Potential glycosylation site"
FT Modified-site 377
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 387
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 398
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 422
FT Modified-site /note= "Potential phosphorylation site"

Db 121 kspqnrnyfyllenawksyiaicfeledsneiftqlyrtlfsvinnghnqkvhnmhvd1 180
QY 181 MSSITCEGDTYSOELLDFVNLVPAHNLNKAQVDAKALLKRTAQALEPVTTFNQV 240
Db 181 mssiccdgtvsqelltdvlnvpahnlknqaydlakalkrtcaqalepyttffnqv 240
QY 241 LMLGKTSISDSSEHFVDFDILELYNDSHLLSVLPQLEFKLSNDNEERLQVVKLARMF 300
Db 241 lmlgktsidsisehvfddilelynidshllsvlpqlfeklsndneerlqvklakmf 300
QY 301 GAKDELSAQNKPLWCQYGRFNDIHVPIRECVKFAFASHCLMNHDPDLAKDITEYLKVRSH 360
Db 301 gakselasnqkplwcylgrfndihvpirlecvkfashclmnhdpdlakditeylkvrsh 360
QY 361 DPEEAIHRDVIIVTAAKDILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKA 420
Db 361 dpeearihdivisvtaakdillvndhllnfvrertldkrwvrkeamglaqiyyka 420
QY 421 LQSAAGKDAAKQIAWKDKLLHIYQNSIDDRLLVERIFAQVMVPHNLETTERMKCLYYL 480
Db 421 lqsaagkaakqiawkdkllhiyqnsidrrllverifagvmvphnlettermkclyy1 480
QY 481 YATDLNAVAKALNEMWKCONLRHQVDLIDLKOPKTDASVKAIFSKVMVITRNLPPDG 540
Db 481 yatdlnavakalnemwkcqnlrhqvkldildikqpktdasvkaifskvmvitrnlppdg 540
QY 541 KAQDPKKFTQVLEDEKIRKOLEVLVSPTCSKQAEQVREIKKLGONKPNQPTNPFLEM 600
Db 541 kaqdfmkftqvledekirkolevlvsptcskqaeqvreitkklgnpkqptnplflem 600
QY 601 TKFLERIAPVHIDPESALIKQVNSIDGTADDEGVPTDQAIRAGLELLKVLSTFH 660
Db 601 tkfleriapvhidpesalikqvnksidgtaddegevgptdqairaglellkvlstfh 660
QY 661 PISFHSATFESLLACLKMDDEKVAEALQIFKNTGSKIEEDFPHIRSALLPVLHKSXK 720
Db 661 pisfhsatfesllaclkmddkvaeealqifkntgskieedfphirsallpvlhksxk 720
QY 721 GPPRAKYAIIHAIKFSKETQFAQIEPLHKSIDPSNLEHLITPLVTIGHIALLAPDQ 780
Db 721 gpprakayaiihaihfsketqfaqieplhksidpsnlehlitplvtighiallapdq 780
QY 781 FAAPKSWVATFIVKDLMLNDRLPKKTKLWVPDVEEVSPTMVKIQAKMVRWLLGMK 840
Db 781 faapkswvatfivkdllmndrlpkktklwvpdeevsptmvmkiqakmvrwllgm 840
QY 841 NNHKSSTGLRLLTILHSDGDLTEQKISKPKDMSRLRLAAGSAIVKLAQEPVHEIIT 900
Db 841 nnhksstglrlltllhsgdgleqkiskpkdmsrlrlaagsaivklagepcyheilit 900
QY 901 LEQYOLCALAINDECYQVROVPAOKLHKLSRLPLEYMAICALCAKDPKERRAHARQ 960
Db 901 leqyolcalaindecyqrvqfagkhglsrlirpleymaicalcakdpkerraharq 960
QY 961 CLVKNINVRREYLKGAHVSEKLLSLLPEYVVPYTHILLAHDPQVYQDIEQLKDVKEC 1020
Db 961 clvkninvrreylkgaavseklslppeyvvythillahdpdyvkqdieqlkdvkec 1020
QY 1021 LMFVLEILMAKNENSHAFIRKWNENIKQTKDAQGDDAKMNEKIYTVCDVAMNIIKSKS 1080
Db 1021 lwfvleilmaknenshafirkwnenikqtkdaqgddakmnekiytcvdvamnliks 1080
QY 1081 TTYTESPKDPVLPARFTQPKNFNTKNYLPPEMKSFETPGPKTKTNVLGAVNKPSS 1140
Db 1081 tctyespkdpvlparrftqpknfntknylppemksfettgpktpktnvlgavnkplss 1140
QY 1141 AGKQSTKSSRMETVSNSSSNPSPGRIGKRLDSSMDHSENEDYTMSSPLPGKKSXK 1200
Db 1141 agkqstksrmetvsnsssnpsprigkrlsdssmdhsenedytmssplpgkksxk 1200
QY 1201 RDDSOLVRSELEKPRGRKTPVTEQEKLGMDDLTKLVQEQPKGSKRKRKGTASBESD 1260
Db 1201 rddsolvrselkprgrktpvteqeklgmddltkvlvqeqpkgsqrsrkrhtasesd 1260

QY 1261 EQQWPEERKRLKEDILENEDEQNSPPKKGRRPPKPLGGTGPKEBPTMKTSKKGSKKSG 1320
Db 1261 eqqweeekrlkdlledenedeqnspkkgrrppkplgggtgpkeptmktksgksg 1320
QY 1321 PPAPEEEEEERQSGNTEQKSKSKOHRYSRRAQRAESPESAIESTQSTPQKGRGRPSK 1380
Db 1321 ppapeeeeeerqsgnteqkskskqhrrysrraqrqraespaiesatstqtpqkgrgrpsk 1380
QY 1381 TPSPSQPKNV 1391
Db 1381 tpspsqpknv 1391

RESULT 2

AAG02811

ID AAG02811 standard; Protein; 92 AA.

XX AAG02811;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 6892.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX N-PSDB; AAC02817.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 6892; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.

XX Sequence 92 AA;

Query Match

Best Local Similarity 6.5%; Score 91; DB 21; Length 92;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 MKNHKSCTSTLRLLTTLILHSDGDLTEQKISKPKDMSRLRLAAGSAIVKLAQEPYHEI 898
|||||

961 6 0.4 16 21 AAY67428
962 6 0.4 16 21 AAY67429
963 6 0.4 16 21 AAY58819
964 6 0.4 16 21 AAY58820
965 6 0.4 16 21 AAY58821
966 6 0.4 17 15 AAR47023
967 6 0.4 17 15 AAR63134
968 6 0.4 17 17 AAW11068
969 6 0.4 17 18 AAW39022
970 6 0.4 17 18 AAW31116
971 6 0.4 17 19 AAW47330
972 6 0.4 17 20 AAY55857
973 6 0.4 18 12 AAR15131
974 6 0.4 18 13 AAR26378
975 6 0.4 18 17 AAW11063
976 6 0.4 18 18 AAW31115
977 6 0.4 18 19 AAW47382
978 6 0.4 18 20 AAY55869
979 6 0.4 18 20 AAY42670
980 6 0.4 18 20 AAY33094
981 6 0.4 19 8 AAF70521
982 6 0.4 19 18 AAW11745
983 6 0.4 20 9 AAP81079
984 6 0.4 20 21 AAY78405
985 6 0.4 20 21 AAY58507
986 6 0.4 20 22 AAB70459
987 6 0.4 20 22 AAB55804
988 6 0.4 20 22 AAB57666
989 6 0.4 20 22 AAB58076
990 6 0.4 21 18 AAW11765
991 6 0.4 21 18 AAW11743
992 6 0.4 22 20 AAY29727
993 6 0.4 23 6 AAP50753
994 6 0.4 24 21 AAB29673
995 6 0.4 24 21 AAY52645
996 6 0.4 25 13 AAR20169
997 6 0.4 25 16 AAR95574
998 6 0.4 25 21 AAR83358
999 6 0.4 26 14 AAR33541
1000 6 0.4 27 18 AAW06240

ALIGNMENTS

RESULT 1
ID AAY94702 standard; Protein; 1391 AA.
XX AAY94702;
AC AAY94702;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human androgen shutoff gene 3 (AS3) protein sequence.
XX
KW Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW diagnosis; treatment; cytostatic; human; ss.
XX
OS Homo sapiens.
XX
FH Key
FT Region Location/Qualifiers
FT 419..459
FT /label= "Subdomain I"
FT /note= "Hank's conserved region"
FT 419..425
FT /label= "Beta strand 1"
FT 426..433
FT /label= "Mg-ATP binding loop"
FT 453..459
FT /label= "Beta strand 2"
FT 472..482
FT /label= "Subdomain II"

Histone H2B peptid
Aggrean H2B peptid
Aggrean degrading
Aggrean degrading
K-ras transformati
Egr-1 nuclear loca
Atherosclerotic pl
Peptide resembling
Atherosclerotic pl
Inhibitor of LDL b
Apolipoprotein fra
Vascular injury af
Sequence of athero
Atherosclerotic pl
Atherosclerotic pl
Inhibitor of LDL b
Apolipoprotein fra
HHV-6 variant B de
Alipoprotein B der
Antigenic oligopep
T-cell receptor pe
Sequence encoded b
Human papillomavir
HHV8 ORF K8.1-deri
L. lactis signal p
PL peptide #15. S
CD105 PL peptide.
CD105 PL peptide.
T-cell receptor pe
T-cell receptor pe
Feline leukaemia v
Sequence of a port
Human/mouse SLE-as
v-ras encoded onco
Fusion junction of
Protein Ki-ras 2A
Conserved d region
Mastoparan analog
Variable heavy cha

FT Region
FT /note= "Hank's conserved domain - beta strand 3"
FT 489..498
FT /label= "Subdomain IIR"
FT /note= "Hank's conserved region - alpha helix C"
FT 509..516
FT /label= "Subdomain IV"
FT /note= "Hank's conserved region - beta strand 4"
FT 525..548
FT /label= "Subdomain V"
FT /note= "Hank's conserved region"
FT 525..532
FT /label= "Beta strand 5"
FT 540..548
FT /label= "Alpha helix D"
FT 554..567
FT /label= "Subdomain VIA"
FT /note= "Hank's conserved region - alpha helix E"
WO200050454-A1.
31-AUG-2000.
24-FEB-2000; 2000WO-US04732.
24-FEB-1999; 99US-0121461.
(TUFT) TUFTS. COLLEGE.
Soto AM, Sonnenschein C, Geck P, Szelel J;
WPI; 2000-565451/52.
N-PSDB; AAA28051, AAA28052.
New human androgen-induced tumor suppressor cDNA sequence termed
'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
diagnosis and treatment of prostate cancer -
Claim 2; Fig 1; 152pp; English.

This invention relates to a human androgen-induced tumour suppressor cDNA
sequence termed 'Androgen Shutoff Gene 3' (AS3). The AS3 gene is located
on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
proliferation and use as a marker for the efficient diagnosis and
treatment of prostate cancer. The invention includes AS3 cDNA and protein
sequences, a vector comprising the cDNA sequence, a host cell transfected
with the expression vector, and a method for producing an AS3 polypeptide
comprising culturing the transfected cells. AS3 has cytostatic activity,
and acts to suppress cell proliferation. The AS3 gene is useful as a
marker for the efficient diagnosis and treatment of prostate cancer. The
AS3 nucleic acid molecule can be used as a source of antisense agents for
sequence specific modulation of gene expression. The AS3 protein may be
used in the treatment of disorders caused by aberrant modification or
mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
or aberrant post-translational modification of the AS3 protein. This
sequence represents the human AS3 protein sequence.

XX Sequence 1391 AA;

Query Match 100.0%; Score 1391; DB 21; Length 1391;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSKTRTNDGKITPPGVKEISDKISKEMVRLKMWKVTMDMDQDSEEEKELYLNLA 60
|||||
DB 1 mahsktrtdgkitppgvkiskiskemvrrlkmvktfmdmdqdsseeekelylnla 60
QY 61 LHLASDFFLKHGKQVRLLVACCLADIFRIYAPEAPYTPSPDKLKIDIFMFTROLKGLDT 120
|||||
DB 61 lhlasdfflkhpqkdvrlvaccladifriyapeapytpdklkdifmfttrlqkldt 120
QY 121 KSPQFNRYFYLLENIAWVKSINICFELSDSNEIFTOLYTLFSVINNGHNQKVMHMDVL 180
|||||

669	7	0.5	649	22	AAB61236	Mature human TANGO	742	7	0.5	859	22	AAB37463	Human ATP-dependen
670	7	0.5	651	20	AAV05243	Human imidazole	743	7	0.5	866	17	AAW04185	Human interleukin-
671	7	0.5	652	20	AAV07094	Colon cancer assoc	744	7	0.5	866	19	AAW61272	Human interleukin-
672	7	0.5	653	20	AAV07050	Renal cancer assoc	745	7	0.5	866	20	AAW99656	Staphylococcus aur
673	7	0.5	654	21	AAV99662	Human GTPase assoc	746	7	0.5	866	20	AAW92409	Human IL-17R prote
674	7	0.5	654	21	AAV41265	Arabidopsis thalia	747	7	0.5	866	21	AAV99941	Human IL-17R prote
675	7	0.5	660	18	AAW18570	Human cyclin D1-hu	748	7	0.5	866	21	AAV97131	Human interleukin-
676	7	0.5	660	19	AAW74571	Human cyclin D1/cy	749	7	0.5	866	21	AAV97181	Human interleukin-
677	7	0.5	662	18	AAW18572	Human cyclin D1-hu	750	7	0.5	866	21	AAW03807	Human interleukin-
678	7	0.5	662	19	AAW4573	Human cyclin D1/cy	751	7	0.5	866	22	AAW62066	Human IL-17R (hCTL
679	7	0.5	664	20	AAV05241	Human imidazole	752	7	0.5	866	22	AAW72754	Human interleukin-
680	7	0.5	664	21	AAV36368	Arabidopsis thalia	753	7	0.5	873	18	AAW19918	Mouse Ksr-1 (kinas
681	7	0.5	667	19	AAW48760	BOP1 protein. Mus	754	7	0.5	875	18	AAW19919	Human Ksr-1 (kinas
682	7	0.5	667	19	AAW57224	Rat proline transp	755	7	0.5	878	18	AAW19113	Potato starch bran
683	7	0.5	667	21	AAV53008	Human secreted pro	756	7	0.5	878	20	AAW97850	Human p101/p13 kin
684	7	0.5	671	21	AAW51790	Arabidopsis thalia	757	7	0.5	879	20	AAW97849	Human p101/p13 kin
685	7	0.5	671	22	AAW61234	Human TANGO 332 pr	758	7	0.5	880	19	AAW23949	Human phosphoinosi
686	7	0.5	672	20	AAW99657	Staphylococcus aur	759	7	0.5	880	20	AAV33720	Human full-length
687	7	0.5	675	17	AAW94303	Human lastin. Hom	760	7	0.5	880	20	AAW97847	Human p101/p13 kin
688	7	0.5	675	21	AAW23361	Human lastin prote	761	7	0.5	880	20	AAW97848	Human p101/p13 kin
689	7	0.5	678	22	AAW30851	Amino acid sequenc	762	7	0.5	880	20	AAW90088	Human G-protein re
690	7	0.5	686	15	AAW60336	Clc-K1 protein. R	763	7	0.5	880	21	AAW28644	Human p101 protein
691	7	0.5	693	22	AAW30845	Amino acid sequenc	764	7	0.5	880	21	AAW08523	Amino acid sequenc
692	7	0.5	705	18	AAW18573	Human cyclin D1-hu	765	7	0.5	880	21	AAV70430	Human phosphatidyl
693	7	0.5	705	19	AAW74574	Human cyclin D1/cy	766	7	0.5	880	21	AAV76800	Human p101 regulat
694	7	0.5	705	21	AAW58193	Lung cancer associ	767	7	0.5	882	17	AAW06399	Class A starch bra
695	7	0.5	706	15	AAW74799	Human APLP2. Homo	768	7	0.5	885	17	AAW99740	Human phosphodiester
696	7	0.5	706	17	AAW98925	Human APLP2. Homo	769	7	0.5	886	9	AAW80345	Sequence encoded b
697	7	0.5	707	16	AAW79912	Human nucleolin.	770	7	0.5	886	20	AAV49828	Human dunce-like p
698	7	0.5	707	20	AAW84052	Human V3 loop HIV	771	7	0.5	886	21	AAW42923	Human ORFX ORF2687
699	7	0.5	707	22	AAW48964	Human nucleolin.	772	7	0.5	886	21	AAW20639	pPDE46 human dunce
700	7	0.5	712	18	AAW30749	Rat YF521 gene pro	773	7	0.5	893	21	AAV96965	Human nuclear dual
701	7	0.5	712	21	AAW56721	Human prostate can	774	7	0.5	896	21	AAV92943	Rat MAGUI 2 prote
702	7	0.5	720	21	AAV94296	Human coenzyme A-u	775	7	0.5	898	20	AAV49804	Human glioblastoma
703	7	0.5	724	22	AAW30850	Amino acid sequenc	776	7	0.5	898	21	AAW20615	pJC44x human glob
704	7	0.5	728	20	AAV43520	Chicken CCTC-bind	777	7	0.5	900	12	AAW14854	Protein associated
705	7	0.5	730	21	AAW39301	Arabidopsis thalia	778	7	0.5	900	17	AAW00091	CAMP phosphodiester
706	7	0.5	730	22	AAW78979	C. glutamicum SRT	779	7	0.5	905	18	AAW31186	Human p160 polyep
707	7	0.5	734	21	AAW20640	pTMS3 human gliobla	780	7	0.5	911	22	AAV97583	Human secreted pro
708	7	0.5	739	19	AAW61532	Mouse Fas-binding	781	7	0.5	912	16	AAW85442	Bovine brevicin co
709	7	0.5	739	21	AAW39300	Arabidopsis thalia	782	7	0.5	919	18	AAW15197	Human EHOC-17 prot
710	7	0.5	739	22	AAW69150	Mouse daxx protein	783	7	0.5	922	21	AAW49845	Arabidopsis thalia
711	7	0.5	739	22	AAW30844	Amino acid sequenc	784	7	0.5	944	20	AAW83318	Mouse SMAD interac
712	7	0.5	740	19	AAW61533	Human Fas-binding	785	7	0.5	956	22	AAW79523	Corynebacterium gl
713	7	0.5	740	22	AAW69149	Human daxx protein	786	7	0.5	963	21	AAV96479	Human vanilloid re
714	7	0.5	742	21	AAW24238	Human vesicle asso	787	7	0.5	968	19	AAW79538	Adenovirus serotyp
715	7	0.5	748	21	AAW42613	Human ORFX ORF2377	788	7	0.5	971	21	AAV98057	Yeast Upflp. Sacc
716	7	0.5	753	20	AAV32132	Death activator de	789	7	0.5	975	19	AAW72746	Drosophila kinesin
717	7	0.5	753	22	AAW30858	Amino acid sequenc	790	7	0.5	979	11	AAW08338	Complete form of h
718	7	0.5	758	20	AAW97677	Human KDS2 protein	791	7	0.5	993	13	AAW27650	Human calcium chan
719	7	0.5	759	21	AAV99487	Yeast acyltransfer	792	7	0.5	999	18	AAW18511	Human oxygen regul
720	7	0.5	763	15	AAW53778	Sequence of human	793	7	0.5	999	18	AAW18512	Rat oxygen regulat
721	7	0.5	776	15	AAW60178	Lethal factor of B	794	7	0.5	1018	17	AAW98747	P. vivax Esp-1 blo
722	7	0.5	787	19	AAW64694	Human ADNF-III pro	795	7	0.5	1018	20	AAW97039	A secreted blood-s
723	7	0.5	789	17	AAW95603	STR4 (suppressor o	796	7	0.5	1032	21	AAV92942	Rat MAGUI 1 prote
724	7	0.5	790	21	AAW49847	Arabidopsis thalia	797	7	0.5	1041	20	AAW30613	Human CDC28-#3 RNA
725	7	0.5	803	20	AAV07113	WO9904265 Seq ID N	798	7	0.5	1045	21	AAW39809	Arabidopsis thalia
726	7	0.5	806	15	AAW56992	hNRP2 U protein.	799	7	0.5	1047	21	AAW41715	Human ORFX ORF1479
727	7	0.5	806	19	AAW64675	Mouse Activity Dep	800	7	0.5	1066	21	AAW39808	Arabidopsis thalia
728	7	0.5	806	21	AAV71135	Mouse Activity Dep	801	7	0.5	1070	20	AAV05244	Human imidazole
729	7	0.5	809	12	AAW15142	Ubiquitin specific	802	7	0.5	1087	20	AAV31146	Human DSIF protein
730	7	0.5	809	14	AAW36729	Ubiquitin-specific	803	7	0.5	1087	22	AAW36645	Human DSIF protein
731	7	0.5	828	11	AAW07661	Incomplete form of	804	7	0.5	1088	18	AAW19786	Human multiple reg
732	7	0.5	828	21	AAV71136	Mouse Activity Dep	805	7	0.5	1098	21	AAW02009	Type VII adenylyl cy
733	7	0.5	829	21	AAW42339	Human ORFX ORF2103	806	7	0.5	1099	21	AAW02011	Type VII adenylyl
734	7	0.5	830	21	AAV96730	PRO539, a Costal-2	807	7	0.5	1103	20	AAV13447	Mouse aortic carbo
735	7	0.5	834	11	AAW08391	Sequence encoded b	808	7	0.5	1112	21	AAW03833	Activin receptor b
736	7	0.5	845	18	AAW36816	Human E2A-binding	809	7	0.5	1123	20	AAV07119	Lung cancer associ
737	7	0.5	846	21	AAV71057	Human membrane tra	810	7	0.5	1128	15	AAW49994	Mouse carboxypepti
738	7	0.5	847	17	AAW06400	Class A starch bra	811	7	0.5	1128	18	AAW36817	Mouse E2A-binding
739	7	0.5	848	19	AAW29677	Homo sapiens clone	812	7	0.5	1130	20	AAV13446	Human aortic carbo
740	7	0.5	848	21	AAW49846	Arabidopsis thalia	813	7	0.5	1132	17	AAW97866	Chicken leucocytos
741	7	0.5	859	20	AAV00991	Human ATP-dependen	814	7	0.5	1135	18	AAW31185	Human p160 polyep

523	7	0.5	414	21	AAG19531	Arabidopsis thalia	596	7	0.5	520	18	AAW15265	Rat beta-amyloid p
524	7	0.5	415	14	AAR32922	AAMP-1. Homo sapi	597	7	0.5	520	21	AAG44273	Arabidopsis thalia
525	7	0.5	416	16	AAR66773	N-terminally trunc	598	7	0.5	523	21	AAG44273	Arabidopsis thalia
526	7	0.5	417	19	AAW47034	Human cartilage gp	599	7	0.5	525	20	AAV14540	Rat lipolysis stim
527	7	0.5	421	19	AAW53121	Amino acid sequenc	600	7	0.5	525	22	AAW59906	Rat lipolysis stim
528	7	0.5	421	19	AAW53122	Amino acid sequenc	601	7	0.5	526	20	AAV14546	Mouse lipolysis st
529	7	0.5	421	19	AAW53122	Amino acid sequenc	602	7	0.5	526	22	AAW59909	Murine leptin frag
530	7	0.5	422	18	AAW11301	Perilipin B (NSEQO	603	7	0.5	528	19	AAW59882	Amino acid sequenc
531	7	0.5	422	21	AAW11301	Perilipin B (NSEQO	604	7	0.5	528	22	AAW61241	Human BFR protein.
532	7	0.5	422	21	AAW11301	Perilipin B (NSEQO	605	7	0.5	530	21	AAW61241	Human BFR protein.
533	7	0.5	422	21	AAW11301	Perilipin B (NSEQO	606	7	0.5	530	21	AAW61241	Human BFR protein.
534	7	0.5	423	21	AAW26392	Arabidopsis thalia	607	7	0.5	531	21	AAW50060	Arabidopsis thalia
535	7	0.5	424	21	AAW26392	Arabidopsis thalia	608	7	0.5	532	21	AAW50059	Arabidopsis thalia
536	7	0.5	425	21	AAW26392	Arabidopsis thalia	609	7	0.5	533	19	AAW50059	Arabidopsis thalia
537	7	0.5	426	16	AAW66776	Cell adhesion prot	610	7	0.5	535	19	AAW61028	Murine guanine nuc
538	7	0.5	426	16	AAW66776	Cell adhesion prot	611	7	0.5	536	21	AAW61028	Murine guanine nuc
539	7	0.5	429	21	AAW28833	Arabidopsis thalia	612	7	0.5	539	21	AAW87747	Human brain proteo
540	7	0.5	429	21	AAW28833	Arabidopsis thalia	613	7	0.5	542	18	AAW36490	Human TULP1 protei
541	7	0.5	430	21	AAW28833	Arabidopsis thalia	614	7	0.5	542	22	AAW28833	Human TULP1 protei
542	7	0.5	431	21	AAW43794	Arabidopsis thalia	615	7	0.5	543	21	AAW43794	Castor bean calnex
543	7	0.5	433	21	AAW39007	Human secreted pro	616	7	0.5	543	21	AAW43794	Arabidopsis thalia
544	7	0.5	433	21	AAW39007	Human secreted pro	617	7	0.5	544	16	AAW82213	Talaromyces flavus
545	7	0.5	433	21	AAW39007	Human secreted pro	618	7	0.5	544	19	AAW82213	Talaromyces flavus
546	7	0.5	435	21	AAW39007	Human secreted pro	619	7	0.5	544	19	AAW82213	Talaromyces flavus
547	7	0.5	438	22	AAW30849	Amino acid sequenc	620	7	0.5	544	21	AAW82213	Talaromyces flavus
548	7	0.5	439	21	AAW30849	Amino acid sequenc	621	7	0.5	547	19	AAW69240	Human doublet. Ho
549	7	0.5	442	22	AAW20330	Human protein phos	622	7	0.5	547	19	AAW69240	Human doublet. Ho
550	7	0.5	442	22	AAW20330	Human protein phos	623	7	0.5	549	21	AAW29189	Arabidopsis thalia
551	7	0.5	444	21	AAW42151	Human ORFX ORF1915	624	7	0.5	552	22	AAW75305	Human secreted pro
552	7	0.5	444	21	AAW42151	Human ORFX ORF1915	625	7	0.5	552	22	AAW75305	Human secreted pro
553	7	0.5	450	18	AAW29454	Streptococcus pneu	626	7	0.5	554	19	AAW61026	Murine guanine nuc
554	7	0.5	450	20	AAW68551	S. pneumoniae Murd	627	7	0.5	559	18	AAW61026	Murine guanine nuc
555	7	0.5	452	21	AAW45369	Arabidopsis thalia	628	7	0.5	560	21	AAW29189	Arabidopsis thalia
556	7	0.5	453	21	AAW47481	Arabidopsis thalia	629	7	0.5	562	16	AAW70491	Leucocytozoan prot
557	7	0.5	455	21	AAW45368	Arabidopsis thalia	630	7	0.5	563	21	AAW16687	Bacteriophage Dp-1
558	7	0.5	456	19	AAW82841	Human cerebral pro	631	7	0.5	568	22	AAW51244	Human calcitonin r
559	7	0.5	456	20	AAW35959	Extended human sec	632	7	0.5	570	18	AAW36002	Amino acid sequenc
560	7	0.5	463	21	AAW67500	Arabidopsis thalia	633	7	0.5	570	21	AAW36002	Amino acid sequenc
561	7	0.5	465	15	AAW50951	PBR-associated pro	634	7	0.5	574	20	AAV14539	Protein encoded by
562	7	0.5	465	18	AAW15266	Receptor for C-ter	635	7	0.5	574	22	AAW59905	Rat lipolysis stim
563	7	0.5	469	22	AAW59875	Beta-amyloid precu	636	7	0.5	575	20	AAV14545	Rat lipolysis stim
564	7	0.5	471	21	AAW90287	Carbon monoxide de	637	7	0.5	575	22	AAW59908	Rat lipolysis stim
565	7	0.5	477	15	AAW66067	Human peptidase, H	638	7	0.5	580	21	AAW29848	Murine leptin frag
566	7	0.5	477	15	AAW66067	Phosphorylated p66	639	7	0.5	580	21	AAW29848	Murine leptin frag
567	7	0.5	477	15	AAW66067	Human OTC. Homo s	640	7	0.5	593	20	AAV14538	Arabidopsis thalia
568	7	0.5	477	20	AAW27591	Human secreted pro	641	7	0.5	593	22	AAW59904	Rat lipolysis stim
569	7	0.5	479	20	AAW97108	Thermostable poly	642	7	0.5	594	14	AAW34936	Rat lipolysis stim
570	7	0.5	479	21	AAW29189	Arabidopsis thalia	643	7	0.5	594	22	AAW34936	Rat lipolysis stim
571	7	0.5	479	22	AAW32389	Herpesvirus entry	644	7	0.5	594	22	AAW34936	Rat lipolysis stim
572	7	0.5	481	14	AAW33389	P. furiosus recomb	645	7	0.5	601	21	AAW79240	Human diacylsulfon
573	7	0.5	481	8	AAW33389	Sequence encoded b	646	7	0.5	605	21	AAW79240	Human diacylsulfon
574	7	0.5	484	22	AAW30848	Amino acid sequenc	647	7	0.5	605	21	AAW79240	Human diacylsulfon
575	7	0.5	487	12	AAW14149	Human alpha 2 beta	648	7	0.5	612	21	AAW49327	Arabidopsis thalia
576	7	0.5	487	18	AAW11804	Human alpha-2b adr	649	7	0.5	612	18	AAW18571	Human cyclin D1-hu
577	7	0.5	489	21	AAW26391	Arabidopsis thalia	650	7	0.5	618	19	AAW474572	Human cyclin D1/cy
578	7	0.5	491	21	AAW22022	Arabidopsis thalia	651	7	0.5	618	21	AAW474572	Human cyclin D1/cy
579	7	0.5	491	21	AAW50056	Arabidopsis thalia	652	7	0.5	619	13	AAW27651	Human calcium chan
580	7	0.5	491	21	AAW50056	Arabidopsis thalia	653	7	0.5	627	21	AAW36369	Arabidopsis thalia
581	7	0.5	493	21	AAW52778	Arabidopsis thalia	654	7	0.5	628	20	AAW37204	Amino acid sequenc
582	7	0.5	495	21	AAW80994	Arabidopsis thalia	655	7	0.5	635	18	AAW19920	Amino acid sequenc
583	7	0.5	496	21	AAW41051	Soybean glycinin A	656	7	0.5	635	19	AAW75855	Human Ksr' (kinase
584	7	0.5	502	21	AAW54580	Human ORFX ORF815	657	7	0.5	635	20	AAW75855	Human secretory pr
585	7	0.5	503	19	AAW53232	A rat calcium/cal	658	7	0.5	635	21	AAW45160	Human secreted pro
586	7	0.5	504	21	AAW52777	Cryptosporidium pa	659	7	0.5	635	21	AAW45160	Human secreted pro
587	7	0.5	505	19	AAW61027	Arabidopsis thalia	660	7	0.5	637	18	AAW11704	High affinity Na+
588	7	0.5	506	21	AAW52776	Murine guanine nuc	661	7	0.5	637	18	AAW11704	High affinity Na+
589	7	0.5	509	21	AAW08254	Arabidopsis thalia	662	7	0.5	638	17	AAW52294	Drosophila p70S6K
590	7	0.5	510	20	AAW29618	Arabidopsis thalia	663	7	0.5	638	17	AAW52294	Drosophila p70S6K
591	7	0.5	517	18	AAW11300	Human PI3K protei	664	7	0.5	642	20	AAW13396	Amino acid sequenc
592	7	0.5	517	18	AAW11300	Human PI3K protei	665	7	0.5	642	21	AAW13396	Amino acid sequenc
593	7	0.5	517	21	AAW32390	Perilipin A (NSEQO	666	7	0.5	642	22	AAW80264	Human PRO332 prote
594	7	0.5	519	21	AAW42468	Rat perilipin A pr	667	7	0.5	646	21	AAW39302	Arabidopsis thalia
595	7	0.5	519	22	AAW61330	Human ORFX ORF2232	668	7	0.5	647	18	AAW00926	Human cyclin D1-hu
						Human transcriptio				647	19	AAW474575	Human cyclin D1/cy
										647	21	AAW93992	Amino acid sequenc

377	7	0.5	296	21	AAG60360	Arabidopsis thalia	450	7	0.5	349	21	AA484617	Amino acid sequenc
378	7	0.5	297	21	AAG07458	Arabidopsis thalia	451	7	0.5	350	21	AAG31413	Arabidopsis thalia
379	7	0.5	298	21	AAG07037	Arabidopsis thalia	452	7	0.5	351	21	AAG29179	Arabidopsis thalia
380	7	0.5	299	21	AAG17712	Arabidopsis thalia	453	7	0.5	352	21	AAG18561	Arabidopsis thalia
381	7	0.5	300	21	AAG36935	Arabidopsis thalia	454	7	0.5	353	21	AAG49090	Arabidopsis thalia
382	7	0.5	301	21	AAG15577	Arabidopsis thalia	455	7	0.5	354	21	AAG30847	Amino acid sequenc
383	7	0.5	302	21	AAG19858	Arabidopsis thalia	456	7	0.5	355	18	AAW14284	Human neuro-D prot
384	7	0.5	303	21	AAG11576	Arabidopsis thalia	457	7	0.5	356	18	AAW22441	Human neurogenic d
385	7	0.5	304	21	AAG25491	Arabidopsis thalia	458	7	0.5	357	19	AAW71017	Human neuroD1 prot
386	7	0.5	305	22	AAB88546	Haemophilus influe	459	7	0.5	358	19	AAW54042	Human BHF1 protein
387	7	0.5	306	21	AAG11575	Arabidopsis thalia	460	7	0.5	359	21	AB114351	Human Neurod 1 pro
388	7	0.5	307	21	AAG50935	Arabidopsis thalia	461	7	0.5	360	21	AB114346	Human Neurod1 prot
389	7	0.5	308	22	AAB46830	A. thaliana TTI pr	462	7	0.5	361	16	AAW77504	Murine neurogenic
390	7	0.5	309	13	AAW20062	Rat EDF-binding pr	463	7	0.5	362	18	AAW22436	Mouse neurogenic d
391	7	0.5	310	21	AAG19857	Arabidopsis thalia	464	7	0.5	363	19	AAW71009	Murine neuroD1 pro
392	7	0.5	311	21	AAW73387	HTRM clone 3340290	465	7	0.5	364	21	AAG18560	Arabidopsis thalia
393	7	0.5	312	22	AAB60476	Human cell cycle a	466	7	0.5	365	21	AAG49089	Arabidopsis thalia
394	7	0.5	313	20	AAG25490	Arabidopsis thalia	467	7	0.5	366	21	AAG05935	Arabidopsis thalia
395	7	0.5	314	21	AAG29339	Arabidopsis thalia	468	7	0.5	367	21	AAG08294	Arabidopsis thalia
396	7	0.5	315	21	AAG19856	Arabidopsis thalia	469	7	0.5	368	17	AAW03627	Human follicle sti
397	7	0.5	316	21	AAG50934	Arabidopsis thalia	470	7	0.5	369	21	AAG37722	Arabidopsis thalia
398	7	0.5	317	21	AAG25489	Arabidopsis thalia	471	7	0.5	370	21	AAG45370	Arabidopsis thalia
399	7	0.5	318	20	AAW34857	Chlamydia pneumoni	472	7	0.5	371	21	AAG11457	Arabidopsis thalia
400	7	0.5	319	21	AAG50933	Arabidopsis thalia	473	7	0.5	372	21	AAG39266	Arabidopsis thalia
401	7	0.5	320	21	AAG47797	Arabidopsis thalia	474	7	0.5	373	19	AAW42427	PDC-E2, BCOADC-E2,
402	7	0.5	321	21	AAW77927	A. thaliana enviro	475	7	0.5	374	21	AAG37789	Arabidopsis thalia
403	7	0.5	322	11	AAW05163	Sequence of human	476	7	0.5	375	21	AAG23057	Arabidopsis thalia
404	7	0.5	323	19	AAW49038	Rabbit low density	477	7	0.5	376	21	AAG18559	Arabidopsis thalia
405	7	0.5	324	21	AAW19760	Human bone sialop	478	7	0.5	377	21	AAG49088	Arabidopsis thalia
406	7	0.5	325	21	AAG47796	Arabidopsis thalia	479	7	0.5	378	21	AAG11456	Arabidopsis thalia
407	7	0.5	326	22	AAW87430	Human gene 11 enco	480	7	0.5	379	21	AAG39265	Arabidopsis thalia
408	7	0.5	327	21	AAG60359	Arabidopsis thalia	481	7	0.5	380	18	AAW25048	BRCA2 cancer suce
409	7	0.5	328	20	AAW85718	Novel protein (C10	482	7	0.5	381	21	AAG24242	Human Nogo B prote
410	7	0.5	329	21	AAG36934	Arabidopsis thalia	483	7	0.5	382	21	AAG36635	Arabidopsis thalia
411	7	0.5	330	21	AAG47795	Arabidopsis thalia	484	7	0.5	383	21	AAW56969	Human MAGI polypep
412	7	0.5	331	21	AAG06457	Arabidopsis thalia	485	7	0.5	384	21	AAW53624	A bone marrow secr
413	7	0.5	332	20	AAW05140	Chlamydia pneumoni	486	7	0.5	385	21	AAW58373	Lung cancer associ
414	7	0.5	333	21	AAG05936	Arabidopsis thalia	487	7	0.5	386	21	AAG34959	Arabidopsis thalia
415	7	0.5	334	21	AAW07097	Colon cancer assoc	488	7	0.5	387	21	AAG06328	Arabidopsis thalia
416	7	0.5	335	21	AAG30241	Arabidopsis thalia	489	7	0.5	388	21	AAW77814	Yeast Upf1 protein
417	7	0.5	336	21	AAG34960	Arabidopsis thalia	490	7	0.5	389	18	AAW22439	Human neurogenic d
418	7	0.5	337	20	AAW35146	Chlamydia pneumoni	491	7	0.5	390	19	AAW71015	Human neuroD2 prot
419	7	0.5	338	21	AAW42303	Human ORFX ORF2067	492	7	0.5	391	18	AAW22442	Mouse neurogenic d
420	7	0.5	339	21	AAW08296	Arabidopsis thalia	493	7	0.5	392	19	AAW71018	Murine neuroD2 pro
421	7	0.5	340	21	AAG18592	Arabidopsis thalia	494	7	0.5	393	21	AAW14347	Human neurogenic b
422	7	0.5	341	21	AAG21956	Arabidopsis thalia	495	7	0.5	394	17	AAW00097	cAMP phosphodieste
423	7	0.5	342	21	AAG37723	Arabidopsis thalia	496	7	0.5	395	20	AAW49817	Human temporal lob
424	7	0.5	343	22	AAW36589	Human FLEXHT-11 pr	497	7	0.5	396	21	AAW20628	PG18ARR human tem
425	7	0.5	344	21	AAW28231	RNA polymerase alp	498	7	0.5	397	17	AAW98521	Xenopus orphan rec
426	7	0.5	345	21	AAW29181	Arabidopsis thalia	499	7	0.5	398	21	AAG36634	Arabidopsis thalia
427	7	0.5	346	21	AAW07036	Arabidopsis thalia	500	7	0.5	399	21	AAG39454	Arabidopsis thalia
428	7	0.5	347	21	AAG21657	Arabidopsis thalia	501	7	0.5	400	21	AAG40293	Arabidopsis thalia
429	7	0.5	348	21	AAG36636	Arabidopsis thalia	502	7	0.5	401	21	AAG30240	Arabidopsis thalia
430	7	0.5	349	21	AAW81647	Streptococcus pneu	503	7	0.5	402	21	AAG35137	Arabidopsis thalia
431	7	0.5	350	21	AAW60983	Arabidopsis thalia	504	7	0.5	403	18	AAW08256	Arabidopsis thalia
432	7	0.5	351	19	AAW27171	HSV-2 strain SB5 C	505	7	0.5	404	21	AAW39264	Lung cancer associ
433	7	0.5	352	19	AAW72090	HSV-2 strain SB5 C	506	7	0.5	405	21	AAW43929	Arabidopsis thalia
434	7	0.5	353	19	AAW2005	HSV-2 strain SB5 C	507	7	0.5	406	21	AAW30846	Arabidopsis thalia
435	7	0.5	354	19	AAW55117	Streptococcus pneu	508	7	0.5	407	21	AAG11455	Arabidopsis thalia
436	7	0.5	355	21	AAW605982	Streptococcus thalia	509	7	0.5	408	21	AAG11455	Arabidopsis thalia
437	7	0.5	356	19	AAW98449	H. pylori GHP0 578	510	7	0.5	409	21	AAW47483	Arabidopsis thalia
438	7	0.5	357	19	AAW79094	Human secreted pro	511	7	0.5	410	21	AAW68011	Yeast immunophilin
439	7	0.5	358	20	AAW49830	Protein sequence S	512	7	0.5	411	21	AAW47482	Human thyrotropin
440	7	0.5	359	20	AAW07754	Human secreted pro	513	7	0.5	412	17	AAW03626	S. frugiperda immu
441	7	0.5	360	21	AAW48519	Bacillus circulans	514	7	0.5	413	20	AAW68010	Human cancer regre
442	7	0.5	361	21	AAW08295	Arabidopsis thalia	515	7	0.5	414	21	AAW09929	Human cancer regre
443	7	0.5	362	21	AAW08295	Arabidopsis thalia	516	7	0.5	415	21	AAG19532	Arabidopsis thalia
444	7	0.5	363	21	AAW53427	Arabidopsis thalia	517	7	0.5	416	22	AAW36604	Human FLEXHT-26 pr
445	7	0.5	364	21	AAW21656	Arabidopsis thalia	518	7	0.5	417	21	AAW36604	Arabidopsis thalia
446	7	0.5	365	18	AAW01561	MOC1 protein, Can	519	7	0.5	418	21	AAW15118	Human FLEXHT-26 pr
447	7	0.5	366	21	AAW15118	Candida cyclin dep	520	7	0.5	419	21	AAW36604	Arabidopsis thalia
448	7	0.5	367	21	AAW42984	Human ORFX ORF2748	521	7	0.5	420	21	AAW36604	Arabidopsis thalia
449	7	0.5	368	19	AAW54365	Human retinitis p1	522	7	0.5	421	21	AAW36604	Arabidopsis thalia

231	7	0.5	200	21	AAG36740	Arabidopsis thalia	304	21	AAG57384	Arabidopsis thalia
232	7	0.5	200	22	AA876848	Human lung tumour	305	21	AA94901	Human secreted pro
233	7	0.5	200	22	AA876868	Human lung tumour	306	21	AA944629	Human myocardium s
234	7	0.5	201	19	AA879540	Adenovirus serotyp	307	22	AA888487	Human membrane or
235	7	0.5	202	21	AAG30242	Arabidopsis thalia	308	22	AA888487	Human secretory pr
236	7	0.5	202	21	AA933263	A dopamine and cam	309	18	AAW09404	Human hepatoma-der
237	7	0.5	204	21	AAG12954	Arabidopsis thalia	310	249	AAG29595	Arabidopsis thalia
238	7	0.5	204	21	AAG38260	Arabidopsis thalia	311	249	AA82156	Human ORFX ORF1920
239	7	0.5	205	21	AAG36739	Arabidopsis thalia	312	252	AA818594	Arabidopsis thalia
240	7	0.5	206	21	AA939598	E. coli L-threonin	313	252	AA821958	Arabidopsis thalia
241	7	0.5	207	21	AA843715	Human cancer assoc	314	252	AA821958	Arabidopsis thalia
242	7	0.5	207	21	AA970950	Mouse oocyte-speci	315	254	AA816726	Arabidopsis thalia
243	7	0.5	207	21	AA970951	Mouse ovary-specif	316	255	AA816717	Arabidopsis thalia
244	7	0.5	207	22	AA863365	Human breast cance	317	255	AA816717	Membrane-bound pro
245	7	0.5	209	21	AA930335	B. subtilis nitro	318	255	AA865240	Human PRO1140 (UNQ
246	7	0.5	209	21	AAG34568	Arabidopsis thalia	319	257	AA831484	Arabidopsis thalia
247	7	0.5	210	21	AAG28047	Arabidopsis thalia	320	258	AA812024	Arabidopsis thalia
248	7	0.5	211	20	AA925422	Human zcalc-1 prot	321	258	AA854905	Arabidopsis thalia
249	7	0.5	212	21	AAG07044	Arabidopsis thalia	322	258	AA860645	Arabidopsis thalia
250	7	0.5	212	21	AAG42682	Arabidopsis thalia	323	259	AA87352	Human gene 11 enco
251	7	0.5	214	21	AA821047	Human nucleic acid	324	261	AA806329	Arabidopsis thalia
252	7	0.5	216	21	AA821047	Hydrophobic domain	325	264	AA843838	Arabidopsis thalia
253	7	0.5	216	21	AA824036	Human PRO407 prot	326	265	AAW42381	Methanococcus igne
254	7	0.5	216	21	AA811736	Arabidopsis thalia	327	265	AA811735	Arabidopsis thalia
255	7	0.5	216	21	AA843331	Arabidopsis thalia	328	265	AA818593	Arabidopsis thalia
256	7	0.5	217	21	AA817714	Arabidopsis thalia	329	265	AA821957	Arabidopsis thalia
257	7	0.5	218	21	AAG04566	Arabidopsis thalia	330	265	AA827773	Arabidopsis thalia
258	7	0.5	218	21	AA859303	Arabidopsis thalia	331	266	AA843330	Arabidopsis thalia
259	7	0.5	219	21	AAG09137	Arabidopsis thalia	332	266	AA816759	Arabidopsis thalia
260	7	0.5	220	21	AA853454	Human colon cancer	333	266	AA823678	Arabidopsis thalia
261	7	0.5	220	21	AA856929	Human prostate can	334	267	AA836738	Arabidopsis thalia
262	7	0.5	221	21	AAG07043	Arabidopsis thalia	335	268	AA838440	Human secreted pro
263	7	0.5	221	21	AA842681	Arabidopsis thalia	336	268	AA800274	Human secreted pro
264	7	0.5	223	21	AA843567	Arabidopsis thalia	337	269	AA843837	Arabidopsis thalia
265	7	0.5	223	21	AA843796	Arabidopsis thalia	338	270	AA827772	Arabidopsis thalia
266	7	0.5	223	21	AA92346	Human cancer assoc	339	271	AA827772	Arabidopsis thalia
267	7	0.5	224	15	AA874004	Tumour rejection a	340	271	AAG23677	Arabidopsis thalia
268	7	0.5	224	16	AA882989	P815A antigen prec	341	271	AAG43795	Arabidopsis thalia
269	7	0.5	224	17	AA894615	MAGE protein. Not	342	273	AAW59646	Amino acid sequenc
270	7	0.5	224	20	AA922127	PLA protein sequen	343	274	AA831414	Arabidopsis thalia
271	7	0.5	224	21	AA87165	Human secreted pro	344	274	AA843836	Arabidopsis thalia
272	7	0.5	224	21	AA876198	Nuclear transport	345	275	AA827771	Arabidopsis thalia
273	7	0.5	225	19	AA885950	S. pneumoniae deri	346	276	AA816757	Arabidopsis thalia
274	7	0.5	225	21	AAG28046	Arabidopsis thalia	347	277	AA850744	Arabidopsis thalia
275	7	0.5	226	21	AA817113	Arabidopsis thalia	348	277	AA814223	Arabidopsis thalia
276	7	0.5	226	21	AA848555	Arabidopsis thalia	349	278	AA812023	Arabidopsis thalia
277	7	0.5	226	21	AA844630	Human mature myoca	350	279	AA860644	Arabidopsis thalia
278	7	0.5	227	18	AAW36148	Rat eosinophil-der	351	281	AAG60643	Signal transductio
279	7	0.5	227	20	AA925425	Human zcalc-1 prot	352	281	AA848050	Arabidopsis thalia
280	7	0.5	227	21	AA858926	Breast and ovarian	353	282	AAG05937	Arabidopsis thalia
281	7	0.5	227	21	AA809136	Arabidopsis thalia	354	282	AA812022	Arabidopsis thalia
282	7	0.5	228	22	AA863361	Human breast cance	355	284	AA871461	Porcine surfactant
283	7	0.5	229	21	AA817250	Arabidopsis thalia	356	284	AA871457	Human prostate tum
284	7	0.5	231	21	AA843925	zeae mays protein f	357	284	AA837724	Arabidopsis thalia
285	7	0.5	231	22	AA820276	Mouse interleukin	358	284	AA95030	Human clone vb22_1
286	7	0.5	232	21	AA825144	Pinus radiata cell	359	285	AA842007	Mouse DNA demethyl
287	7	0.5	234	13	AA829717	Guinea pig MBP-2.	360	285	AA841938	Human ORFX ORF1702
288	7	0.5	234	19	AAW40076	Guinea pig eosinop	361	286	AA813505	Arabidopsis thalia
289	7	0.5	234	21	AA831415	Arabidopsis thalia	362	287	AA813504	Arabidopsis thalia
290	7	0.5	234	22	AA820278	Mouse interleukin	363	287	AA821332	Human ORFX ORF1896
291	7	0.5	236	15	AA856668	Bacteroides fragil	364	288	AA832752	Short-chain tumour
292	7	0.5	240	20	AAW92295	Mouse alpha-1 (XVI	365	291	AA814198	Human DNA demethyl
293	7	0.5	240	21	AAG07042	Arabidopsis thalia	366	292	AA812056	Arabidopsis thalia
294	7	0.5	240	21	AA842680	Arabidopsis thalia	367	292	AA842996	Arabidopsis thalia
295	7	0.5	241	21	AAG29981	Arabidopsis thalia	368	293	AA81938	Mouse rad17 cell c
296	7	0.5	245	21	AAG38221	Arabidopsis thalia	369	293	AA813503	Arabidopsis thalia
297	7	0.5	246	21	AA857385	Arabidopsis thalia	370	294	AA860361	Arabidopsis thalia
298	7	0.5	247	21	AAG29980	Arabidopsis thalia	371	295	AA827114	pradi. Homo sapie
299	7	0.5	247	21	AA857952	Human transmembran	372	295	AA844802	Human cyclin D1.
300	7	0.5	248	20	AA825421	Human zcalc-1 alie	373	295	AA854044	bcl-1 protein. Ho
301	7	0.5	248	21	AA818922	A novel polypeptid	374	295	AA8606459	Arabidopsis thalia
302	7	0.5	248	21	AA824034	Human PRO4354 prot	375	295	AA820848	Arabidopsis thalia
303	7	0.5	248	21	AAG10894	Arabidopsis thalia	376	295	AA829340	Arabidopsis thalia
									AA851130	Arabidopsis thalia

85	7	0.5	16	15	AAR66081	Phosphorylated db1	158	7	0.5	123	21	AAG12897	Arabidopsis thalia
86	7	0.5	16	15	AAR66085	Phosphorylated db1	159	7	0.5	124	22	AAB45728	Human 7TM clone H7
87	7	0.5	16	15	AAR66069	Phosphorylated db1	160	7	0.5	126	21	AAB41594	Human OREX ORF1358
88	7	0.5	16	15	AAR66076	Phosphorylated db1	161	7	0.5	128	21	AAG04003	Human secreted pro
89	7	0.5	20	15	AAR57412	Peptide for treati	162	7	0.5	128	21	AAW87070	Human secreted pro
90	7	0.5	22	14	AAR31230	Prepro-thyrotropin	163	7	0.5	129	21	AAG33107	Arabidopsis thalia
91	7	0.5	25	20	AAV36550	Fragment of human	164	7	0.5	129	21	AAG36780	Arabidopsis thalia
92	7	0.5	25	21	AAV67504	PBR-associated pro	165	7	0.5	129	21	AAG01626	Human secreted pro
93	7	0.5	37	18	AAW45249	Peptide conjugate	166	7	0.5	131	19	AAW58851	Human AC222.1 secr
94	7	0.5	45	22	AAB80386	Secreted protein e	167	7	0.5	131	19	AAW37144	Human AC222.1 prot
95	7	0.5	51	19	AAW79542	Adenovirus serotyp	168	7	0.5	131	21	AAG12896	Arabidopsis thalia
96	7	0.5	53	21	AAV53800	Amino acids 145-19	169	7	0.5	131	21	AAG33106	Arabidopsis thalia
97	7	0.5	58	21	AAG23408	Arabidopsis thalia	170	7	0.5	131	21	AAG36779	Arabidopsis thalia
98	7	0.5	58	21	AAG41160	Zea mays protein f	171	7	0.5	131	21	AAV70954	Arabidopsis SERK i
99	7	0.5	64	21	AAG05618	Arabidopsis thalia	172	7	0.5	134	21	AAG02224	Human secreted pro
100	7	0.5	64	21	AAG56760	Arabidopsis thalia	173	7	0.5	139	18	AAW27864	Staphylococcus aur
101	7	0.5	65	21	AAB52182	Human secreted pro	174	7	0.5	139	21	AAG12854	Arabidopsis thalia
102	7	0.5	68	21	AAG14241	Arabidopsis thalia	175	7	0.5	141	21	AAG08433	Arabidopsis thalia
103	7	0.5	68	21	AAG02223	Human secreted pro	176	7	0.5	141	21	AAG12853	Arabidopsis thalia
104	7	0.5	72	19	AAV21092	Human p53 cellular	177	7	0.5	141	21	AAG54057	Arabidopsis thalia
105	7	0.5	72	21	AAG10911	Arabidopsis thalia	178	7	0.5	142	18	AAW55361	H. pylori ORF 04ce
106	7	0.5	73	21	AAG00403	Human secreted pro	179	7	0.5	143	19	AAW38734	Streptococcus pneu
107	7	0.5	75	21	AAG05617	Arabidopsis thalia	180	7	0.5	143	20	AAV12496	Human 5' EST secre
108	7	0.5	75	21	AAG23802	Arabidopsis thalia	181	7	0.5	143	21	AAV68981	Cps9H protein whic
109	7	0.5	75	21	AAG01779	Human secreted pro	182	7	0.5	145	21	AAG12895	Arabidopsis thalia
110	7	0.5	77	21	AAG05616	Arabidopsis thalia	183	7	0.5	147	19	AAW72404	Arabidopsis thalia
111	7	0.5	79	21	AAB57237	Human prostate can	184	7	0.5	148	21	AAG12852	Arabidopsis thalia
112	7	0.5	81	21	AAG15127	Arabidopsis thalia	185	7	0.5	148	21	AAG23800	Arabidopsis thalia
113	7	0.5	81	21	AAG37763	Arabidopsis thalia	186	7	0.5	152	21	AAB19764	Human bone sialopr
114	7	0.5	81	21	AAG37763	Arabidopsis thalia	187	7	0.5	153	21	AAG37408	Arabidopsis thalia
115	7	0.5	83	21	AAG23801	Arabidopsis thalia	188	7	0.5	153	21	AAG55089	Arabidopsis thalia
116	7	0.5	85	21	AAB54302	Human pancreatic c	189	7	0.5	154	20	AAV35957	Extended human sec
117	7	0.5	85	21	AAG57367	Arabidopsis thalia	190	7	0.5	154	20	AAV16786	Human secreted pro
118	7	0.5	89	21	AAG05819	Arabidopsis thalia	191	7	0.5	155	19	AAV72382	Arabidopsis pathog
119	7	0.5	90	21	AAV53798	Amino acids 145-23	192	7	0.5	156	16	AAV77506	Human neurogenic d
120	7	0.5	91	21	AAG10910	Arabidopsis thalia	193	7	0.5	156	18	AAW22438	Human neurogenic d
121	7	0.5	91	21	AAG03640	Human secreted pro	194	7	0.5	156	19	AAW71014	Partial human neur
122	7	0.5	92	21	AAG02036	Human secreted pro	195	7	0.5	156	21	AAB40971	Human OREX ORF735
123	7	0.5	95	21	AAG06923	Arabidopsis thalia	196	7	0.5	157	20	AAV60331	Human normal bladd
124	7	0.5	95	21	AAG17942	Arabidopsis thalia	197	7	0.5	157	21	AAB53376	Human colon cancer
125	7	0.5	95	21	AAG36553	Arabidopsis thalia	198	7	0.5	158	18	AAW55514	H. pylori ORF 05ee
126	7	0.5	96	21	AAG49201	Arabidopsis thalia	199	7	0.5	159	21	AAG54056	Arabidopsis thalia
127	7	0.5	98	21	AAB471941	Arabidopsis thalia	200	7	0.5	161	16	AAV77507	Human neurogenic d
128	7	0.5	98	22	AAB47130	CDiff-9, Incyte ID	201	7	0.5	162	21	AAG29982	Arabidopsis thalia
129	7	0.5	99	21	AAG32968	Arabidopsis thalia	202	7	0.5	164	22	AAB63364	Human breast cance
130	7	0.5	101	20	AAV36549	Fragment of human	203	7	0.5	166	21	AAG01567	Human secreted pro
131	7	0.5	101	21	AAB58979	Breast and ovarian	204	7	0.5	168	21	AAG12701	Arabidopsis thalia
132	7	0.5	105	21	AAB56522	Human prostate can	205	7	0.5	170	21	AAG12700	Arabidopsis thalia
133	7	0.5	105	21	AAG15126	Arabidopsis thalia	206	7	0.5	171	20	AAV37804	Protein which is s
134	7	0.5	105	21	AAG36862	Arabidopsis thalia	207	7	0.5	172	21	AAG43926	Zea mays protein f
135	7	0.5	105	21	AAG37759	Arabidopsis thalia	208	7	0.5	174	21	AAG12956	Arabidopsis thalia
136	7	0.5	105	21	AAG37762	Arabidopsis thalia	209	7	0.5	174	21	AAG38262	Arabidopsis thalia
137	7	0.5	106	21	AAG06026	Arabidopsis thalia	210	7	0.5	176	21	AAG12955	Arabidopsis thalia
138	7	0.5	107	21	AAG05818	Arabidopsis thalia	211	7	0.5	176	21	AAG38261	Arabidopsis thalia
139	7	0.5	107	8	AAG06025	Arabidopsis thalia	212	7	0.5	177	15	AAAG0904	Borrelia IP90 anti
140	7	0.5	109	8	AAV71074	Human prothymosin	213	7	0.5	177	15	AAAG60905	Borrelia NSB1AB an
141	7	0.5	109	9	AAV82976	Human pro-thymosin	214	7	0.5	177	15	AAAG62790	Borrelia IP90 anti
142	7	0.5	109	12	AAR13680	Thymosin alphas fr	215	7	0.5	177	15	AAAG62791	Borrelia NSB1AB an
143	7	0.5	109	21	AAG03806	Human secreted pro	216	7	0.5	177	21	AAG12699	Arabidopsis thalia
144	7	0.5	110	21	AAV37876	Human prothymosin	217	7	0.5	177	21	AAG33761	Arabidopsis thalia
145	7	0.5	114	21	AAG01672	Human secreted pro	218	7	0.5	178	21	AAG29185	Arabidopsis thalia
146	7	0.5	115	21	AAG32967	Arabidopsis thalia	219	7	0.5	178	21	AAG34230	Arabidopsis thalia
147	7	0.5	115	21	AAV28420	Corn branched chai	220	7	0.5	179	21	AAG36936	Arabidopsis thalia
148	7	0.5	116	21	AAG49200	Arabidopsis thalia	221	7	0.5	180	13	AAG22392	Antigen tc-7a, Ei
149	7	0.5	117	20	AAV32780	HIV chemokine gene	222	7	0.5	181	21	AAG28048	Arabidopsis thalia
150	7	0.5	118	19	AAW42424	Branched chain 2-o	223	7	0.5	181	21	AAG44516	Arabidopsis thalia
151	7	0.5	120	21	AAB53238	Human colon cancer	224	7	0.5	182	21	AAG59751	Arabidopsis thalia
152	7	0.5	120	21	AAG15125	Arabidopsis thalia	225	7	0.5	184	16	AAW67278	wheat ubiquitin-co
153	7	0.5	120	21	AAG36861	Arabidopsis thalia	226	7	0.5	188	21	AAG20849	Arabidopsis thalia
154	7	0.5	120	21	AAG37758	Arabidopsis thalia	227	7	0.5	188	21	AAG51131	Arabidopsis thalia
155	7	0.5	120	21	AAG37761	Arabidopsis thalia	228	7	0.5	189	21	AAW41686	Human OREX ORF1450
156	7	0.5	121	21	AAG54940	Arabidopsis thalia	229	7	0.5	190	20	AAV41399	Human secreted pro
157	7	0.5	121	22	AAB87644	Bovine mammary tis	230	7	0.5	198	21	AAV95939	Porcine adenovirus

THIS PAGE BLANK (USPTO)

Search completed: September 25, 2001, 22:57:52
Job time: 23890 sec

Human; P2; CX5C chemokine; Chromosome 5q31; gene therapy; asthma;
allergic rhinitis; urticaria; anaphylactic shock; hives; hay fever; ds

Sequence 1067 BP; 331 A; 170 C; 197 G; 369 T; 0 other;

Best local similarity 100.00, freq. no. 10,
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID AAF18145 standard: DNA: 1216 BP.

AA
AC

DT 14-MAR-2001 (first entry)

Lung cancer associated polynucleotide sequence SEQ ID 164.

KW proliferative disorder: wound healing: infectious disease: ds.

OS Homo sapiens.

PN WO200055180-A2.

PD 21-SEP-2000.

08-MAR-2000; 2000WO-US05918.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC

XX
XX

XXXXXX

DR P-PSDB; AAB58269.

Lung cancer assoc

PT such as lung cancer -

PS Claim 1; Page 627-628

CC The invention relates to the isolation of genes AAC78997-C79045 encoding
CC 49 human secreted proteins AAB44335-B44382. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC78988) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC the proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

Db 76 acatttttttttttttttttgg 101

RESULT 34
AAV89475/C
ID AAV89475 standard; DNA; 293 BP.
XX
AC AAV89475;
XX
13-APR-1995 (first entry)
XX
Junction A of SfiI fragment contg. Newcastle Disease Virus genes.
DB XX
KW SfiI fragment; Newcastle Disease Virus; junction; ss.
XX
OS Synthetic.
XX
Key Location/Qualifiers
FH misc_feature 1..39 /*tag= a
FT /label= linker
FT promoter 40..78 /*tag= b
FT /label= EP1
FT promoter 79..121 /*tag= c
FT /label= LP2
FT misc_feature 122..182 /*tag= d
FT /label= FRAGMENT 1, NDV
XX
W09419014-A.
PN XX
XX
01-SRP-1994.
PD XX
XX
28-FEB-1994; 94WO-USO1826.
PF XX
XX
26-FEB-1993; 93US-0024156.
PR XX
XX
(JAPG) NIPPON ZEON KK.
PA PA
(SYTR) SYNTRO CORP.
PI XX
Cochran MD;
PS XX
WPI: 1994-294007/36.
XX
New recombinant fowl pox virus for use in vaccines - contains
PT genes expressing antigens of Newcastle disease virus and opt.
PT infectious bronchitis virus
PT
PT Disclosure; Figure 1; 85pp; English.
PS
XX
Newcastle Disease Virus (NDV) HN and F genes were inserted as a
CC SfiI fragment into the homology vector 443-88.8 at the unique
CC SfiI site. The NDV HN and F genes were inserted in the same
CC transcriptional orientation as the ORF in the parental homology
CC vector. The sequence of SfiI fragment is in AAQ68943/R58858/AA58859.
CC The inserted SfiI fragment has the following structure:
CC Junction A - Fragment 1 (HN, AAs 2-577) - Junction B - Fragment
CC 2 (F, AAs 1-553) - Junction C - Fragment 3 (PBR322) - Junction D.
CC Fragment 1 is approx. 1811 bp Avail to NaeI fragment of the NDV HN
CC cDNA clone (B1 strain). Fragment 2 is an approx 1812 bp BamHI to
CC PstI fragment of the full length NDV F cDNA (B1 Strain). Fragment
CC 3 is an approx 235 bp PstI and ScaI fragment of the plasmid pBR322.
CC The sequences of the Junctions are in AAQ68945, AAQ68946, AAQ68947 and
CC AAQ68948.

Query Match 0.5%; Score 26; DB 15; Length 182;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
SQ Sequence 182 BP; 52 A; 33 C; 42 G; 55 T; 0 other;

XX
RESULT 35
AAT88088/C
ID AAT88088 standard; cDNA; 394 BP.
XX
AC AAT88088;

XX
SQ Sequence 293 BP; 93 A; 68 C; 52 G; 80 T; 0 other;

Query Match 0.5%; Score 26; DB 20; Length 293;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4434 tcttttttttttttttttttttttg 4459
|||||
Db 90 TCTTTTCTTTTCTTTTCTTTTCTTTTG 65

XX
RESULT 35
AAT88088/C
ID AAT88088 standard; cDNA; 394 BP.
XX
AC AAT88088;


```

FT primer_bind /*tag= as
FT complement (6772..6790)
FT /*tag= at
FT /note= "downstream amplification primer 10-504"
FT 7120..7137
FT /*tag= au
FT /note= "upstream amplification primer 10-204"
FT 7422..7468
FT /*tag= av
FT /note= "potential binding site for a probe"
FT replace (7445, A)
FT /*tag= aw
FT /note= "claim 4"
FT 7513..7531
FT /*tag= ax
FT /note= "upstream amplification primer 10-32"
FT complement (7557..7574)
FT /*tag= ay
FT /note= "downstream amplification primer 10-204"
FT 7612..7637
FT /*tag= az
FT /note= "specifically claimed in claim 3"
FT 7709..7852
FT /*tag= ba
FT /number= 1
FT 7783..78290
FT /*tag= bb
FT /product= FLAP
FT /note= "contains introns"
FT 7847..7893
FT /*tag= bc
FT /note= "potential binding site for a probe"
FT 7853..16235
FT /*tag= bd
FT /number= 1
FT replace (7870, A)
FT /*tag= be
FT /note= "claim 4"
FT complement (7914..7933)
FT /*tag= bf
FT /note= "downstream amplification primer 10-32"
FT 8117..15994
FT /*tag= bg
FT /note= "specifically claimed in claim 1"
FT 16114..16132
FT /*tag= bh
FT /note= "upstream amplification primer 10-33"
FT 16236..16335
FT /*tag= bi
FT /number= 2
FT 16265..16311
FT /*tag= bj
FT /note= "potential binding site for a probe"
FT replace (16288, T)
FT /*tag= bk
FT /note= "claim 4"
FT 16324..16370
FT /*tag= bl
FT /note= "potential binding site for a probe"
FT 16336..24226
FT /*tag= bm
FT /number= 2
FT replace (16347, A)
FT /*tag= bn
FT 16360..16406

```

```

Query Match 0.5%; Score 27; DB 21; Length 43069;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 4432 aatcttttttttttttttttttttttgg 4458
Db 29028 AATCTTTTTTTTTTTTTTTTTTTTGTG 29002

```

RESULT 32

AAD00579

ID AAD00579 standard; DNA; 128 BP.

XX AAD00579;

XX AAD00579;

XX 21-SEP-2000 (first entry)

XX Human Hscdc6 intron 2 of genomic sequence.

Hscdc6; CSH; human; DNA replication; mitosis; cyclin-dependent kinase; cdk; regulator; inhibitor; cdc6p; cdc18; vertebrate cdc6; antagonist; agonist; treatment; cell proliferative disease; psoriasis; medicament; atherosclerotic vascular disease; vascular restenosis; cardiomyopathy; inflammatory arthritis; autoimmune disease; organ transplant rejection; traumatic injury; stroke; myocardial infarction; renal failure; rejection; hepatic failure; cancer; breast; colon; cervix; lymphoma; diagnosis; screen; cytostatic; immunosuppressive; antiinflammatory; antiarthritic; antipsoriatic; antiarteriosclerosis; vasotropic; cardiac; vulnery; tranquiliser; cerebroprotective; intron; ds.

Homo sapiens.

OS WO2000026242-A2.

XX 11-MAY-2000.

XX 28-OCT-1999; 99WO-US25445.

XX 30-OCT-1998; 98US-0183266.

XX (COLD-) COLD SPRING HARBOR LAB.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Stillman B, Williams RS, Mendez JR;

XX WPI; 2000-365574/31.

XX DNA replication-regulating gene, which is vertebrate cdc6 gene and its antagonist is useful for treating cell proliferative diseases such as atherosclerotic vascular disease, vascular restenosis, psoriasis

XX Claim 5; Fig 8B; 11pp; English.

The present DNA sequence is the intron 2 of human Hscdc6 genomic sequence (previously referred as CSH), involved in cell cycle regulation. Hscdc6 functions as a regulator of DNA replication and/or entry of a cell into mitosis and as a cyclin-dependent kinase (cdk) inhibitor. It is expressed in active proliferative cells. It has similarity to cdc18 and cdc6p proteins from Saccharomyces cerevisiae and S. pombe, respectively. The agonist or antagonist of vertebrate cdc6 gene is useful for treating cell proliferative diseases like atherosclerotic vascular disease, vascular restenosis, psoriasis, inflammatory arthritis, autoimmune diseases and organ transplant rejection. Cdc6 gene or agonist is useful for treating conditions associated with loss of viable tissues, such as traumatic injury, stroke, myocardial infarction, cardiomyopathy, renal and hepatic failure, by enhancing cell proliferation. An antagonist comprising cdc6 specific antibody is useful to manufacture medicament for treatment of cancer of the breast, colon, cervix or lymphoma. The antibodies are used to screen for hscdc6 and diagnosis of a proliferative disorder.

XX Sequence 128 BP; 24 A; 22 C; 19 G; 63 T; 0 other;

Query Match 0.5%; Score 26; DB 21; Length 128;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4434 tcttttttttttttttttttttttttgg 4459

Db 103 tcttttttttttttttttttttttttgg 128

Query Match 0.5%; Score 27; DB 22; Length 28690;
 Best Local Similarity 100.0%; Pred. No. 4.2;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4433 atctttttttttttttttttttttttttttttgg 4459

DB 22470 atctttttttttttttttttttttttttttttgg 22496

RESULT 31

AAZ36335/C

ID AAZ36335 standard; DNA; 43069 BP.

XX AAZ36335;

XX 22-FEB-2000 (first entry)

XX Genomic sequence of the 5-lipoxygenase activating protein (FLAP).

XX Human; 5-lipoxygenase activating protein; FLAP; biallelic marker;
 KW leukotriene pathway; genotype: haplotype; FLAP-related biallelic marker;
 KW asthma; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT misc_feature 1..7007

FT /tag= a

FT /note= "specifically claimed in claim 1"

FT misc_feature 1..7708

FT /tag= b

FT /note= "potential 5' regulatory region"

FT primer_bind 3851..3869

FT /tag= c

FT /note= "upstream amplification primer 10-517"

FT allele replace (3950, C)

FT primer_bind 3927..3973

FT /tag= e

FT /note= "potential binding site for a probe"

FT primer_bind complement (4171..4189)

FT /tag= f

FT /note= "downstream amplification primer 10-517"

FT primer_bind 4120..4138

FT /tag= g

FT /note= "upstream amplification primer 10-518"

FT primer_bind 4220..4266

FT /tag= h

FT /note= "potential binding site for a probe"

FT allele replace (4243, T)

FT primer_bind 4289..4335

FT /tag= j

FT /note= "potential binding site for a probe"

FT allele replace (4312, G)

FT primer_bind complement (4372..4390)

FT /tag= k

FT /note= "downstream amplification primer 10-518"

FT primer_bind 4373..4391

FT /tag= m

FT /note= "upstream amplification primer 10-253"

FT primer_bind 4467..4513

FT /tag= n

FT /note= "potential binding site for a probe"

FT allele replace (4490, G)

FT primer_bind 4647..4693

FT /tag= o

FT /note= "potential binding site for a probe"

FT primer_bind 4664..4710

FT /tag= p

FT /note= "potential binding site for a probe"

FT primer_bind complement (4711..4719)

FT /tag= q

FT /note= "potential binding site for a probe"

FT allele replace (4670, G)

FT /tag= r

FT allele replace (4687, C)

FT /tag= s

FT primer_bind complement (4773..4792)

FT /tag= t

FT primer_bind /note= "downstream amplification primer 10-253"

FT 4814..4833

FT /tag= u

FT /note= "upstream amplification primer 10-499"

FT 4945..4991

FT /tag= v

FT /note= "potential binding site for a probe"

FT 4956..4972

FT /tag= w

FT /note= "upstream amplification primer 10-500"

FT allele replace (4968, A)

FT /tag= x

FT primer_bind complement (5026..5043)

FT /tag= y

FT /note= "downstream amplification primer 10-499"

FT 5117..5163

FT /tag= z

FT /note= "potential binding site for a probe"

FT allele replace (5140, T)

FT /tag= aa

FT primer_bind 5190..5236

FT /tag= ab

FT /note= "potential binding site for a probe"

FT allele replace (5213, G)

FT /tag= ac

FT primer_bind 5341..5387

FT /tag= ad

FT /note= "potential binding site for a probe"

FT allele replace (5364, A)

FT /tag= ae

FT primer_bind complement (5405..5422)

FT /tag= af

FT /note= "downstream amplification primer 10-500"

FT 5524..5542

FT /tag= ag

FT /note= "upstream amplification primer 10-522"

FT 5571..5617

FT /tag= ah

FT /note= "potential binding site for a probe"

FT allele replace (5594, A)

FT /tag= ai

FT primer_bind complement (5978..5996)

FT /tag= aj

FT /note= "downstream amplification primer 10-522"

FT 6218..6235

FT /tag= ak

FT /note= "upstream amplification primer 10-503"

FT 6347..6393

FT /tag= al

FT /note= "potential binding site for a probe"

FT allele replace (6370, G)

FT /tag= am

FT primer_bind 6522..6539

FT /tag= an

FT /note= "upstream amplification primer 10-504"

FT 6670..6716

FT /tag= ao

FT /note= "downstream amplification primer 10-503"

FT primer_bind 6670..6716

FT /tag= ap

FT /note= "potential binding site for a probe"

FT allele replace (6693, T)

FT /tag= aq

FT primer_bind 6740..6786

FT /tag= ar

FT /note= "potential binding site for a probe"

FT allele replace (6763, A)

XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX WPI: 1999-551358/46.
XX P-PSDB; AAV41764.
XX New secreted and transmembrane polypeptides and their polynucleotides,
XX useful for treating blood coagulation disorders, cancers and cellular
XX adhesion disorders -
XX Claim 2; Fig 201; 530pp; English.
XX The present invention describes secreted and transmembrane polypeptides
XX and their polynucleotides. The nucleotide sequences are useful as
XX sources of probes, primers, for chromosome mapping, and for generation
XX of antisense sequences. They can also be used to create transgenic
XX animals. The proteins can be used to treat a variety of diseases and
XX disorders, depending on their function. Diseases that may be treated
XX include blood coagulation disorders, cancers and cellular adhesion
XX disorders. They may also be used to raise antibodies. AA233891 to
XX AA234338, and AA41685 to AA41774 represent polynucleotide and
XX polypeptide sequence given in the exemplification of the present
XX invention.
XX Sequence 2477 BP; 724 A; 478 C; 495 G; 780 T; 0 other;
SQ

Query Match 0.5%; Score 27; DB 20; Length 2477;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4432 aatctttttttttttttttttttttttttttttg 4458
|||||
Db 1668 aatctttttttttttttttttttttttttttttg 1694

RESULT 27
AAC78579
ID AAC78579 standard; cDNA; 2477 BP.
XX AAC78579;
XX 08-FEB-2001 (first entry)
XX Human PRO1082 (UN0539) nucleotide sequence SEQ ID NO:476.
XX Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
XX expressed sequence tag; detection; cancer; ss.
XX Homo sapiens.
XX WO200053756-A2.
XX 14-SEP-2000.
XX 18-FEB-2000; 2000WO-US04341.
XX 08-MAR-1999; 99WO-US05028.
XX 12-MAR-1999; 99US-0123957.
XX 29-MAR-1999; 99US-0126773.
XX 21-APR-1999; 99US-0130232.
XX 28-APR-1999; 99US-0131445.
XX 14-MAY-1999; 99US-0134287.
XX 23-JUN-1999; 99US-0141037.
XX 26-JUL-1999; 99US-0145698.
XX 29-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 02-DEC-1999; 99WO-US28565.
XX 16-DEC-1999; 99WO-US30095.
XX 30-DEC-1999; 99WO-US31243.
XX 30-DEC-1999; 99WO-US31274.
XX 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00277.
XX 06-JAN-2000; 2000WO-US00376.
XX (GETH) GENENTECH INC.
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
XX Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
XX Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX WPI: 2000-611443/58.
XX P-PSDB; AAB44320.
XX Novel PRO polypeptides and polynucleotides used in detection methods,
XX to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX Claim 2; Fig 201; 636pp; English.
XX WPI: 2000-611443/58.
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
XX sequence tag) sequences which encode secreted or transmembrane PRO
XX polypeptides. The PRO polynucleotides and polypeptides have cytostatic
XX activity. The polynucleotides and polypeptides can be used for detecting
XX the presence of PRO polypeptides in samples, for linking bioactive
XX molecules to cells and for modulating biological activities of cells,
XX using the polypeptides for specific targeting. The polypeptide targeting
XX can be used to kill the target cells, e.g. for the treatment of cancers.
XX The polypeptide pairs provide specific targeting of bioactive molecules
XX to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
XX the isolation of the PRO polynucleotide sequences.
XX Sequence 2477 BP; 724 A; 478 C; 495 G; 780 T; 0 other;
SQ

Query Match 0.5%; Score 27; DB 21; Length 2477;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4432 aatctttttttttttttttttttttttttttttg 4458
|||||
Db 1668 aatctttttttttttttttttttttttttttttg 1694

RESULT 28
AAX02987
ID AAX02987 standard; DNA; 3076 BP.
XX AAX02987;
XX 22-JUN-1999 (first entry)
XX Human IL-lra BAC contiguous DNA sequence 32.
XX Tango-77; human; IL-lra; cytokine superfamily; inflammation; inhibition;
XX interleukin-1 receptor; IL-lr; regulation; asthma; rheumatoid arthritis;
XX chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
XX growth factors; treatment; IL-1 receptor complex; BAC; ss.
XX Homo sapiens.
XX WO9906426-A1.
XX 11-FEB-1999.
XX 03-AUG-1998; 98WO-US16102.
XX 02-JUL-1998; 98US-0091650.
XX 04-AUG-1997; 97US-0054646.
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX Pan Y;


```

PR 18-DEC-1997;      97US-0068054.
PR 18-DEC-1997;      97US-0068057.
PR 18-DEC-1997;      97US-0068064.
PR 18-DEC-1997;      97US-0070923.
PR 18-DEC-1997;      97US-0068169.
PR 18-DEC-1997;      97US-0068365.
PR 18-DEC-1997;      97US-0068367.
PR 18-DEC-1997;      97US-0068368.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;
PI Florence K, Greene JM, Janat F, Kyaw H, Moore PA;
PI Ni J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;
PI Yu G;
XX
XX WPI; 1999-418749/35.
DR P-PSDB; AAY36310.
DR
XX
XX New isolated human genes encoding secreted polypeptides
PT
XX Claim 1; Page 324; 537pp; English.
PS
XX
CC AAX97916 to AAX98029 represent 110 isolated human secreted protein
CC genes. AAY36224 to AAY36727 represent the secreted proteins encoded by
CC the 110 human genes. The genes and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new genes.
CC Specific uses are described for each of the 110 genes, based on which
CC tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours, developmental
CC abnormalities and foetal deficiencies, blood disorders, diseases of the
CC immune system, autoimmune diseases, inflammation, allergies, Alzheimer's
CC and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis,
CC sepsis, skin disorders, atherosclerosis, diabetes, cardiovascular
CC disorders, kidney disorders, digestive/endocrine disorders, infections
CC and AIDS. The polypeptides are also useful for identifying their binding
CC partners. The sequences given in AAX97907 to AAX97915 and AAY36223 are
CC used in the exemplification of the present invention.
XX
SQ Sequence 1352 BP; 474 A; 259 C; 248 G; 371 T; 0 other;

Query Match          0.5%; Score 27; DB 20; Length 1352;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches   27; Conservative 0; Mismatches    0; Indels    0; Gaps    0

QY 4431 aaactctttttttttttttttttt 4457
Db 317 aaactctttttttttttttttttt 343
|||||
RESULT 23
AAC59286
ID AAC59286 standard; cDNA; 1358 BP.
XX
AC AAC59286;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein cDNA #10.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotrophic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW candiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
PN WQ2000056753-A1.

```


DR WPI; 2000-594630/56.
 XX P-PSDB; AAB39350.
 XX
 PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; Page 343; 395pp; English.
 XX
 CC The polynucleotide sequences given in AAC74280 to AAC74327 encode the
 CC human secreted proteins given in AAB39310 to AAB39357. AAB39358 to
 CC AAB39400 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC antithrombotic; immunosuppressive; antirheumatic; antiproliferative;
 CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
 CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They can also be used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases, hyperproliferative disorders, cardiovascular
 CC disorders, cerebrovascular disorders, angiogenesis, nervous system
 CC disorders. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 680 BP; 208 A; 128 C; 109 G; 235 T; 0 other;

Query Match 0.5%; Score 27; DB 21; Length 680;
 Best Local Similarity 100.0%; Pred. No. 9.9;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4434 tctttttttttttttttttttttgt 4460
 Db | ttttttttttttttttttttttttttttttt
 653 TCTTTTCTTTTCTTTTCTTTTCTTTTGT 627
 RESULT 20
 AAX40124
 ID AAX40124 standard; DNA; 704 BP.
 XX
 AC AAX40124;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Gastric cancer associated gene.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9904265-A2.
 XX
 PD 28-JAN-1999.
 XX
 PF 15-JUL-1998; 98WO-US14679.
 XX
 XX 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR

PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX
 PA Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX
 XX WPI; 1999-132448/11.
 DR
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 PS Claim 67; Page 705; 787pp; English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SQ Sequence 704 BP; 142 A; 144 C; 197 G; 202 T; 19 other;
 Query Match 0.5%; Score 27; DB 20; Length 704;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4434 tctttttttttttttttttttttgt 4460
 Db | ttttttttttttttttttttttttttttttt
 19 tctttttttttttttttttttttttttttgt 45
 RESULT 21
 AAC59318/c
 ID AAC59318 standard; cDNA; 845 BP.
 XX
 AC AAC59318;
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE Human secreted protein cDNA #42.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200056753-A1.
 XX
 PD 28-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-US06765.
 XX
 XX 23-MAR-1999; 99US-0126051.
 PR 10-DEC-1999; 99US-0169906.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.


```
XX SQ Sequence 530 BP; 191 A; 90 C; 83 G; 166 T; 0 other;
Query Match 1.8%; Score 97; DB 21; Length 530;
Best Local Similarity 100.0%; Pred. No. 8.3e-21;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 172 agatgggttgaaacatttatgatatggaccaggactctgaagaagaaaggagcttt 231
DQ 215 AGATGGTGTGAAACATTTATGATATGGACAGGACTCTGAAGAAGAAAGAGCTTT 156
QY 232 atttaaacctagctttacatcttgcctcagattttt 268
DQ 155 ATTTAAACCTAGCTTTACATCTTGCTTCAGATTTT 119
RESULT 13
AAAX21198
ID AAX21198 standard; DNA; 421 BP.
AC AAX21198;
AT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
XX 23-JUN-1998; 98WO-US13041.
XX 24-JUN-1997; 97US-0050667.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Fraser CM;
XX WPI; 1999-081273/07.
XX New isolated Treponema pallidum nucleic acids - used to develop
XX products for the detection, diagnosis, characterisation, prevention
XX and therapy of T. pallidum infections, particularly syphilis
XX Claim 1; Page 1119; 1150pp; English.
XX AAX20500-21243 represent polynucleotide sequences from the genome of
XX Treponema pallidum. The sequences can be used for detection,
XX diagnosis, characterisation, prevention and therapy for T. pallidum
XX infections, particularly syphilis. They can also be used for detecting
XX diseases related to Borrelia infections in animals, and for the
XX production of biosynthetic products such as enzymes.
XX SQ Sequence 421 BP; 96 A; 86 C; 100 G; 139 T; 0 other;
Query Match 0.6%; Score 29; DB 20; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4429 aaaaatcttttttttttttttttttttttttttttt 4457
DQ 188 aaaaatcttttttttttttttttttttttttttttt 216
RESULT 14
AAV87322
ID AAV87322 standard; CDNA; 489 BP.
AC AAV87322;
AT 27-APR-1999 (first entry)
DE EST clone BP810.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; ESF; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845435-A2.
PD 15-OCT-1998.
XX 10-APR-1998; 98WO-US06954.
XX 10-APR-1997; 97US-0835913.
XX (CEMY ) GENETICS INST INC.
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX Racie LA, Spaulding V, Treacy M;
XX WPI; 1999-070076/06.
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries
XX Claim 1; Page 527-528; 633pp; English.
XX This sequence represents an expressed sequence tag (EST), and is a
XX polynucleotide of the invention. The polynucleotides of the invention are
XX all secreted EST sequences isolated from a variety of human tissue
XX sources. The EST sequences and proteins encoded by them are predicted to
XX have useful biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals, although no supporting data is given. Suggested activities
XX include nutritional activity, immune stimulating or suppressing activity,
XX haematopoiesis regulating activity, tissue growth activity, haemostatic
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, cadherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The EST sequences are also stated to be useful for gene
XX therapy.
XX SQ Sequence 489 BP; 159 A; 95 C; 116 G; 119 T; 0 other;
Query Match 0.6%; Score 29; DB 20; Length 489;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4429 aaaaatcttttttttttttttttttttttttttttt 4457
DQ 139 aaaaatcttttttttttttttttttttttttttttt 167
RESULT 15
AAA64141/c
ID AAA64141 standard; DNA; 14784 BP.
AC AAA64141;
AT 20-DEC-2000 (first entry)
DE Nucleotide sequence of a beta-tubulin antigen.
XX
```


Sequence 150 BP; 43 A; 26 C; 38 G; 42 T; 1 other;
Query Match 1.9%; Score 99; DB 19; Length 150;
Best Local Similarity 99.3%; Pred. No. 2.8e-21;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4491 cacaaatgggactgctgaagagtgacaggttgacatttggacccatacatt 4550
Db 1 cacaaatgggactgctgaagagtgacaggttgacatttggacccatacatt 60
QY 4551 tttgtgcacatctttagccatacacatgtaacattgacttgagtgcttggaaagt 4610
Db 61 tttgtgcacatctttagccatacacatgtaacattgacttgagtgcttggaaagt 120
QY 4611 taatgtgcagtgctatgtagacataaaga 4640
Db 121 taatgtgcagtgctatgtagacataaaga 150
RESULT 11
AAZ12050
ID AAZ12050 standard; DNA; 150 BP.
XX AC AAZ12050;
XX DT 30-MAR-1999 (first entry)
XX DE Human biallelic polymorphic DNA fragment WI-18387a.
XX KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
XX OS Homo sapiens.
XX PN WO9820165-A2.
XX PD 14-MAY-1998.
XX PF 05-NOV-1997; 97WO-US20313.
XX PR 06-NOV-1996; 96US-0030455.
XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX PI Hudson T, Lander ES, Wang D;
XX DR WPI; 1998-286974/25.
XX PS New isolated nucleic acid segments from the human genome - used for
XX determining polymorphic forms for use in e.g. forensics, paternity
XX testing or phenotypic typing for disease
XX Claim 1; Page 216; 310pp; English.
XX CC AAX10269-X12937 are human DNA fragments which contain biallelic
XX polymorphic markers which have been isolated using the primers
XX represented in AAX09121-X10268. The base occupying the polymorphic site
XX is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
XX can be used in methods for determining polymorphic forms in an individual
XX for use in e.g. forensics, paternity testing or for phenotypic typing for
XX diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
XX syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
XX familial hypercholesterolemia, polycystic kidney disease, hereditary
XX spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
XX haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
XX syndrome, osteogenesis imperfecta, acute intermittent porphyria,
XX autoimmune diseases, inflammation, cancer, diseases of the nervous
XX system, infection by pathogenic microorganisms, and characteristics such
XX as longevity, appearance (e.g. baldness, obesity), strength, speed,
XX endurance, fertility, and susceptibility or receptivity to particular

drugs or therapeutic treatments. The isolated polymorphic nucleic acid
segments can also be used to produce medicaments for the treatment or
prophylaxis of such diseases.
Sequence 150 BP; 43 A; 26 C; 38 G; 42 T; 1 other;
Query Match 1.9%; Score 99; DB 19; Length 150;
Best Local Similarity 99.3%; Pred. No. 2.8e-21;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4491 cacaaatgggactgctgaagagtgacaggttgacatttggacccatacatt 4550
Db 1 cacaaatgggactgctgaagagtgacaggttgacatttggacccatacatt 60
QY 4551 tttgtgcacatctttagccatacacatgtaacattgacttgagtgcttggaaagt 4610
Db 61 tttgtgcacatctttagccatacacatgtaacattgacttgagtgcttggaaagt 120
QY 4611 taatgtgcagtgctatgtagacataaaga 4640
Db 121 taatgtgcagtgctatgtagacataaaga 150
RESULT 12
AAZ80598/c
ID AAZ80598 standard; cDNA; 530 BP.
XX AC AAZ80598;
XX DT 07-APR-2000 (first entry)
XX DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:682.
XX KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX OS Homo sapiens.
XX PN WO9964576-A2.
XX PD 16-DEC-1999.
XX PF 09-JUN-1999; 99WO-IB01062.
XX PR 10-JUN-1998; 98US-0088801.
XX PA (FARB) BAYER CORP.
XX PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX DR WPI; 2000-087220/07.
XX PT Novel nucleic acids, used to develop products for the diagnosis and
XX treatment of disorders involving unwanted cell proliferation,
XX particularly cancers, especially colon cancer
XX Claim 15; Page 402; 469pp; English.
XX CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
XX the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
XX cDNA clones can be used to generate antisense oligonucleotides which
XX can be used for antisense therapy. Methods and products from the present
XX invention can be used for identifying and/or classifying cancerous cells
XX present in a human tumour, particularly in solid tumours, e.g.
XX carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
XX can be used for developing agents for the diagnosis and treatment of
XX disorders involving unwanted cell proliferation, such as neoplasia,
XX dysplasia or hyperplasia.

CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.

XX Sequence 405 BP; 99 A; 63 C; 76 G; 146 T; 21 other;

Query Match 2.5%; Score 131; DB 16; Length 405;
 Best Local Similarity 100.0%; Pred. NO. 5e-31;
 Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4912 cccttctgtatcttaccatagtttactctctggtggacccttaattctcagagtgctaa 4971
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 132 cccttctgtatcttaccatagtttactctctggtggacccttaattctcagagtgctaa 191
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 4972 attctgtccattacaccagagatgctctctgatagaggaacaccatgcaaatgtga 5031
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 192 attctgtccattacaccagagatgctctctgatagaggaacaccatgcaaatgtga 251
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 5032 aatagtctctga 5042
 ||||||||||

Db 252 aatagtctctga 262

RESULT 9

AAC06611 ID AAC06611 standard; cDNA; 161 BP.

AC AAC06611;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 10686.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 10686; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX

SQ Sequence 161 BP; 54 A; 25 C; 43 G; 39 T; 0 other;

Query Match 2.3%; Score 120; DB 21; Length 161;
 Best Local Similarity 100.0%; Pred. NO. 1.3e-27;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 45 aggggtagaataatttctgtcattcgaagactagaccatgatggaaaatt 104
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 42 aggggtagaataatttctgtcattcgaagactagaccatgatggaaaatt 101
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 105 acatctcgcctgggttcaaggaataatcagataataatctaaagagagatggtgaga 164
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 102 acatctcgcctgggttcaaggaataatcagataataatctaaagagagatggtgaga 161
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 10

AAX12049 ID AAX12049 standard; DNA; 150 BP.

XX AAX12049;

XX 30-MAR-1999 (first entry)

XX Human biallelic polymorphic DNA fragment WI-18387b.

XX Polymorphism; biallelic; human; forensic; paternity testing; disease;
 KW detection; phenotypic typing; characteristic; infection; hereditary;
 KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
 KW treatment; marker; ss.

XX Homo sapiens.

XX WO9820165-A2.

XX 14-MAY-1998.

XX 05-NOV-1997; 97WO-US20313.

XX 06-NOV-1996; 96US-0030455.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Hudson T, Lander ES, Wang D;

XX WPI; 1998-286974/25.

XX New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics, paternity
 PT testing or phenotypic typing for disease

PS Claim 1; Page 215; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain biallelic
 CC polymorphic markers which have been isolated using the primers
 CC represented in AAX09121-X10268. The base occupying the polymorphic site
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
 CC can be used in methods for determining polymorphic forms in an individual
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
 CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.

XX

QY 347 agctccttacacatccct 365
 Db 328 agctccttacacatccct 346

RESULT 7

AAZ14459
 ID AAZ14459 standard; cDNA; 284 BP.

AC AAZ14459;

XX 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:1928.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

XX Claim 1; Page 1096; 2479pp; English.

CC The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ1779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

XX Sequence 284 BP; 89 A; 43 C; 46 G; 103 T; 3 other;

Query Match 4.2%; Score 219; DB 20; Length 284;

Best Local Similarity 100.0%; Pred. No. 1.6e-57;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4970 aaattgtctgccattacaccagaagatgctctgtatagaggacacacattgt 5029
 Db 1 aaattgtctgccattacaccagaagatgctctgtatagaggacacacattgt 60

QY 5030 gaaatagctcctgaagttcttgattacttacacccagattgttgcaccagaattt 5089
 Db 61 gaaatagctcctgaagttcttgattacttacacccagattgttgcaccagaattt 120

QY 5090 tctggcctttcatggcaatgaaatttttaagaagaagatttaaaagtattttatttaa 5149
 Db 121 tctggcctttcatggcaatgaaatttttaagaagaagatttaaaagtattttatttaa 180

QY 5150 agagtgtgtataaaataatgtactgaattctttatccc 5188
 Db 181 agagtgtgtataaaataatgtactgaattctttatccc 219

RESULT 8

AAT19926

ID AAT19926 standard; cDNA to mRNA; 405 BP.

XX AC AAT19926;

XX 17-JUL-1996 (first entry)

XX Human gene signature HUMGS01054.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.

XX Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATS/) MATSUBARA K.

XX (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

XX Claim 1; Page 510-511; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridize to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS

PF 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG02811.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 2815; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

XX Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

Query Match 5.5%; Score 292; DB 21; Length 295;

Best Local Similarity 100.0%; Pred. No. 1.6e-79;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2562 gtctgagtcactggaataaataacacagtaataacaggaactctaccttaaga 2621

Db 1 gtctgagtcactggaataaataacacagtaataacaggaactctaccttaaga 60

QY 2622 ttgtcaacaacaattatcatagtgagagacttgacagacaggggaaattagtaaa 2681

Db 61 ttgtcaacaacaattatcatagtgagagacttgacagacaggggaaattagtaaa 120

QY 2682 ccagatatgtcacgtctgagactgtctgtgagtgctattgtgaagctggcacaagaa 2741

Db 121 ccagatatgtcacgtctgagactgtctgtgagtgctattgtgaagctggcacaagaa 180

QY 2742 cccgtttaccatgaataatcacattagaaacaataatcagctatgtgattagctatcaac 2801

Db 181 cccgtttaccatgaataatcacattagaaacaataatcagctatgtgattagctatcaac 240

QY 2802 gatagaatctatcaagtaagaacaagtgttgcacagaaacttcacaaaggcc 2853

Db 241 gatagaatctatcaagtaagaacaagtgttgcacagaaacttcacaaaggcc 292

RESULT 6

AAZ42861

ID AAZ42861 standard; cDNA; 351 BP.

XX AC AAZ42861;

XX 01-FEB-2000 (first entry)

XX Human 5' EST isolated from a cDNA library SEQ ID NO:620.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;

regulation; identification; ss.

XX Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

XX 09-APR-1999; 99WO-IB00712.

XX 09-APR-1998; 98US-0057719.

XX 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-038446/03.

XX P-PSDB; AAY65247.

XX Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

XX Claim 1; Page 495; 837pp; English.

XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
 CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 351 BP; 107 A; 63 C; 80 G; 99 T; 2 other;

Query Match 5.1%; Score 268; DB 21; Length 351;

Best Local Similarity 99.7%; Pred. No. 2.6e-72;

Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 47 gggtagaataatttctgctcatggctcattcaagaactaggaccaatgatgaaaaattac 106

Db 28 gggtagaataatttctgctcatggctcattcaagaactaggaccaatgatgaaaaattac 87

QY 107 atatccgctgggtccaaggaaataatcagataaaatatactaaaggagagtggtgagacg 166

Db 88 atatccgctgggtccaaggaaataatcagataaaatatactaaaggagagtggtgagacg 147

QY 167 attaaagatggttgaaacttttcatgatatgaccagactctgaagaagaaagga 226

Db 148 attaaagatggttgaaacttttcatgatatgaccagactctgaagaagaaagga 207

QY 227 gctttattaaacctagctttacatcttgccttcagatttttttccaagcatcctggttaa 286

Db 208 gctttattaaacctagctttacatcttgccttcagatttttttccaagcatcctggttaa 267

QY 287 agatgttcgtactggtgagctgctgctgctgctgctgctgctgctgctgctgctgctg 346

Db 268 agatgttcgtactggtgagctgctgctgctgctgctgctgctgctgctgctgctgctg 327


```
QY 3290 catcatgtcaagagtagtactacatacagtttggaattctctaaagaccggtagtaccagc 3349
DB 22 catcatgtcaagagtagtactacatacagtttggaattctctaaagaccggtagtaccagc 81
QY 3350 tcgtttcttcactcaactgacaagaatttcagtaacacacaaattatctgctcctga 3409
DB 82 tcgtttcttcactcaactgacaagaatttcagtaacacacaaattatctgctcctga 141
QY 3410 atgaataatcattttcaactcctggaaacctaaacacaaatgttcttaggagctgttaa 3469
DB 142 atgaataatcattttcaactcctggaaacctaaacacaaatgttcttaggagctgttaa 201
QY 3470 caagccacttcatcagcaggaagaactctcagaccacaaatcagcaatggaactgt 3529
DB 202 caagccacttcatcagcaggaagaactctcagaccacaaatcagcaatggaactgt 261
QY 3530 aagcaatgcaagcagcagctcaaatccaagctctcctggaagaataaaaggagagctga 3589
DB 262 aagcaatgcaagcagcagctcaaatccaagctctcctggaagaataaaaggagagctga 321
QY 3590 tagttctgaaatggatcacagtgaaatgaagattacacaatgtcttcacctttgcccgg 3649
DB 322 tagttctgaaatggatcacagtgaaatgaagattacacaatgtcttcacctttgcccgg 381
QY 3650 gaaaaaaagtgaagagagagagactctgatctttgtaaggtctgaat 3696
DB 382 gaaaaaaagtgaagagagagagactctgatctttgtaaggtctgaat 428
```

RESULT 4

AAZ15259
ID AAZ15259 standard; cDNA; 738 BP.

AAZ15259;

12-OCT-1999 (first entry)

Human gene expression product cDNA sequence SEQ ID NO:2728.

Human; gene; gene expression product; diagnosis; therapy; probe;

detection; mapping; tissue typing; profiling; forensic; cancer;

genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

Homo sapiens.

WO9338372-A2.

05-AUG-1999.

28-JAN-1999; 99WO-US01619.

03-APR-1998; 98US-0080666.

28-JAN-1998; 98US-0072910.

24-FEB-1998; 98US-0075954.

31-MAR-1998; 98US-0080114.

03-APR-1998; 98US-0080515.

(CHIR) CHIRON CORP.

(HYSE-) HYSEQ INC.

Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

Lanson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

Stache-Crain B, Sudduth-Klinger J, Williams LT;

WPI; 1999-494092/41.

Novel human genes and their expression products which are

differentially expressed in different cell types

XX

PS

Claim 1; Page 1324; 2479pp; English.

XX

CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ15259 to AAZ1779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ15259 to AAZ1779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.

SQ Sequence 738 BP; 268 A; 145 C; 168 G; 150 T; 7 other;

Query Match 6.6%; Score 348; DB 20; Length 738;

Best Local Similarity 100.0%; Pred. No. 1.7e-96;

Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3890 agaaaaatgaagatgaacagaatagtcgcccaaaaaagggttaaaagagcgccaccacaaa 3949

DB 51 agaaaaatgaagatgaacagaatagtcgcccaaaaaagggttaaaagagcgccaccacaaa 110

QY 3950 acctcttggtgaggtacaccacaaagaagagccaaatgaaactctcaaaaaaggag 4009

DB 111 acctcttggtgaggtacaccacaaagaagagccaaatgaaactctcaaaaaaggag 170

QY 4010 caaaaaaatctggacctccagccagagagagagaggaagaagaacaaagtgg 4069

DB 171 caaaaaaatctggacctccagccagagagagagaggaagaagaacaaagtgg 230

QY 4070 aaatcaggaacagaagtcacaaacacacagcagcagagtgctcaagagagcagcagag 4129

DB 231 aaatcaggaacagaagtcacaaacacacagcagcagagtgctcaagagagcagcagag 290

QY 4130 agcagaatctcctgaatctagtgcaattgaattccacacagtcacacccacagaaagcag 4189

DB 291 agcagaatctcctgaatctagtgcaattgaattccacacagtcacacccacagaaagcag 350

QY 4190 aggaagaccatcaaaaaagccatccatccatcaacacacacacacacacacacacacac 4237

DB 351 aggaagaccatcaaaaaagccatccatccatccatcaacacacacacacacacacacacacac 398

RESULT 5

AAC02817

ID AAC02817 standard; cDNA; 295 BP.

AAC02817;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 2815.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

XX

[illegible]

Qy	5101	atgccaatgaataattttaagaagaagatttaaagtatttttaattttaagaagtgtgttta	5160
Db	5185	atgccaatgaataattttaagaagaagatttaaagtatttttaattttaagaagtgtgttta	5244
Qy	5161	taaaataatgtactgaattctttatccccattttatccattttatcatcttccagtttttatttaatactta	5220
Db	5245	taaaataatgtactgaattctttatccccattttatccattttatcatcttccagtttttatttaatactta	5304
Qy	5221	ctgtatcaataaaaattctgttaatttgaatgagtaaaaaaaataaaaaa	5271
Db	5305	ctgtatcaataaaaattctgttaatttgaatgagtaaaaaaaataaaaaa	5355
RESULT	3		
AAV87629			
ID	AAV87629	standard; cDNA; 439 BP.	
XX			
AC	AAV87629;		
XX			
DT	12-FEB-1999	(first entry)	
XX			
DE	EST clone DY17.		
XX			
KW	Expressed sequence tag; secreted protein; haematopoiesis regulator;		
KW	tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;		
KW	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;		
KW	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO9845437-A2.		
XX			
PD	15-OCT-1998.		
XX			
PF	10-APR-1998; 98WO-US06956.		
XX			
PR	10-APR-1997; 97US-0837312.		
XX			
PA	(GEMY) GENETICS INST INC.		
XX			
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;		
PI	Racie LA, Spaulding V, Treacy M;		
XX			
XX	WPI; 1999-070078/06.		
DR			
XX			
PT	New polynucleotides encoding human secreted proteins - derived from		
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,		
PT	ovary, pituitary, retina and colon cDNA libraries		
XX			
PS	Claim 1; Page 126; 641pp; English.		
XX			
CC	The present sequence represents an expressed sequence tag (EST), and is		
CC	a polynucleotide of the invention. The polynucleotides of the invention		
CC	are all secreted EST sequences isolated from a variety of human tissue		
CC	sources. The EST sequences and proteins encoded by them are predicted to		
CC	have useful biological activities which would make them suitable for		
CC	treating, preventing or ameliorating medical conditions in humans and		
CC	animals, although no supporting data is given. Suggested activities		
CC	include nutritional activity, immune stimulating or suppressing activity		
CC	haematopoiesis regulating activity, tissue growth activity,		
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory		
CC	activity, cadherin/tumour invasion suppressor activity, tumour inhibition		
CC	therapy. The EST sequences are also stated to be useful for gene		
XX			
SQ	Sequence 439 BP; 150 A; 101 C; 85 G; 103 T; 0 other;		

Query Match 7.7%; Score 407; DB 20; Length 439;
Best Local Similarity 100.0%; Pred. No. 3.2e-114;
Matches 407; Conservative 0; Mismatches 0; Indels

1885 gfgtgcgtgaaataactaagaagttgggcaaccccaacagcgtacaaatccttctcgtg 1944
1861 aaatgatcaagttctcttgagagagatagcacctctgacacatagatcacgaatctatca 1920
1945 aaatgatcaagttctcttgagagagatagcacctctgacacatagatcacgaatctatca 2004
1921 gfgctcttataaacaagtgaacaaatcaatagatggaacagacagatgatgaagtggag 1980
2005 gfgctcttataaacaagtgaacaaatcaatagatggaacagacagatgatgaagtggag 2064
1981 gfgtcccaactgatcaagccatcaagcaggcttgaaactgcttaaggtactctcaattta 2040
2065 gfgtcccaactgatcaagccatcaagcaggcttgaaactgcttaaggtactctcaattta 2124
2041 cacatcccactctcattcttctgtaaacatttgaactcattactggtctgctgaaaa 2100
2125 cacatcccactctcattcttctgtaaacatttgaactcattactggtctgctgaaaa 2184
2101 tggatgatgaaaaagttagcagaagctgcactacaaaatttttcaaaaaacacaggaagcaaa 2160
2185 tggatgatgaaaaagttagcagaagctgcactacaaaatttttcaaaaaacacaggaagcaaa 2244
2161 ttgaagagatattccacacatcaagatcagccttgcttctgctttttacatcacaaatctca 2220
2245 ttgaagagatattccacacatcaagatcagccttgcttctgctttttacatcacaaatctca 2304
2221 aaaaagaccccccgctcaagccaaataagccatttattgattccatcgatatttttctca 2280
2305 aaaaagaccccccgctcaagccaaataagccatttattgattccatcgatatttttctca 2364
2281 gtaaaagagaccagtttgacagatatttgagcctctgcataagagccctagatccaaagca 2340
2365 gtaaaagagaccagtttgacagatatttgagcctctgcataagagccctagatccaaagca 2424
2341 acctggaacatctcataaacacattggttactatttggtcatattgctctccttgcaacctg 2400
2425 acctggaacatctcataaacacattggttactatttggtcatattgctctccttgcaacctg 2484
2401 atcaatttgcctctcttgaaatcttggttagctacttttcaatttggaagatcttctca 2460
2485 atcaatttgcctctcttgaaatcttggttagctacttttcaatttggaagatcttctca 2544
2461 tgaatgatcggttccagggaagaaagacaaactaaacttgggttccagatgaagaagtat 2520
2545 tgaatgatcggttccagggaagaaagacaaactaaacttgggttccagatgaagaagtat 2604
2521 ctccctgagacaatggtcaaaaattcaggctattaaaaatgatggttcgatggtacttggaa 2580
2605 ctccctgagacaatggtcaaaaattcaggctattaaaaatgatggttcgatggtacttggaa 2664
2581 tgaaaaaataatcacagtaaatcaggaaacttctaccttaagattggttaacaaacaatttgc 2640
2665 tgaaaaaataatcacagtaaatcaggaaacttctaccttaagattggttaacaaacaatttgc 2724
2641 atagtgatggagacttgacagaaacagggaataatagtaaacagatatgtcacgtctga 2700
2725 atagtgatggagacttgacagaaacagggaataatagtaaacagatatgtcacgtctga 2784
2701 gacttgcctggtgagtgctatttgtaagctggcacagaacccctgttaccatgaatatca 2760
2785 gacttgcctggtgagtgctatttgtaagctggcacagaacccctgttaccatgaatatca 2844
2761 tcacattagaacaataatcagctatgtgcatatagctatcaacgatgaatgctatcaagtaa 2820
2845 tcacattagaacaataatcagctatgtgcatatagctatcaacgatgaatgctatcaagtaa 2904
2821 gacaagtgttgcccaagaacttcaacaagggcttccggtttacggcttccacttgagt 2880
2905 gacaagtgttgcccaagaacttcaacaagggcttccggtttacggcttccacttgagt 2964
2881 atatggcaatctgtgccttctgcaaaagatcctctgtaagagagagagctcatgcta 2940

2965 atatggcaatctgtgccccttctgtcaaaaagatcctctgaaaggagagaagctcatgcta 3024
2941 ggcaatgtttggtgaaaaataaaaatgtaaggcgaggatctatctgaaagcagcatgcagctg 3000
3025 ggcaatgtttggtgaaaaataaaaatgtaaggcgaggatctatctgaaagcagcatgcagctg 3084
3001 ttatgtgaaaaattattgtctcttctaccagagatattgttccatatacaaatcaccttt 3060
3085 ttatgtgaaaaattattgtctcttctaccagagatattgttccatatacaaatcaccttt 3144
3061 ttgacacatgacccagatattgtcaaaagtacagatattgaacaacttaagatgtttaaag 3120
3145 ttgacacatgacccagatattgtcaaaagtacagatattgaacaacttaagatgtttaaag 3204
3121 aatgtctttggttctgtgaaaataattaatggctaaaaatgaaataaacagtcacgctt 3180
3205 aatgtctttggttctgtgaaaataattaatggctaaaaatgaaataaacagtcacgctt 3264
3181 ttatcagaagaatggttagaaaattattaaacaacaaagaatgcccaagacagatgatg 3240
3265 ttatcagaagaatggttagaaaattattaaacaacaaagaatgcccaagacagatgatg 3324
3241 caaaaatgaatgaaaaactgtacactgtgtgtgatttgcctatgaatatcatcatc 3300
3325 caaaaatgaatgaaaaactgtacactgtgtgtgatttgcctatgaatatcatcatc 3384
3301 agagtactacatcacagtttggaaatcctctaaagacccggttactaccagctcgtttctca 3360
3385 agagtactacatcacagtttggaaatcctctaaagacccggttactaccagctcgtttctca 3444
3361 ctcaacctgacaagaatttcagtaacacccaaaaattatctgctcctgaaatgaatcat 3420
3445 ctcaacctgacaagaatttcagtaacacccaaaaattatctgctcctgaaatgaatcat 3504
3421 ttttcaactctgaaaaacccctaaacaacaaatgttctagagcgtttaaacaagccactt 3480
3505 ttttcaactctgaaaaacccctaaacaacaaatgttctagagcgtttaaacaagccactt 3564
3481 catcagcagcagcaatctcagaccaaactcatcagaatggaactgtaagcaatgcaaa 3540
3565 catcagcagcagcaatctcagaccaaactcatcagaatggaactgtaagcaatgcaaa 3624
3541 ggagcagctcacaatccaactcctctggaagaataaaagggggagcgttgatgttctgaaa 3600
3625 ggagcagctcacaatccaactcctctggaagaataaaagggggagcgttgatgttctgaaa 3684
3601 ttgattcacagtgaataatgaagattcacaaatgtcttcaacttgcgagggaataaagt 3660
3685 ttgattcacagtgaataatgaagattcacaaatgtcttcaacttgcgagggaataaagt 3744
3661 acaagagagacgactctgatcttctgaaagctggaattggagaagcctagaggcagaaaa 3720
3745 acaagagagacgactctgatcttctgaaagctggaattggagaagcctagaggcagaaaa 3804
3721 aaacgcccgtcacagaacaggagggaataataggatggatgacttgacttaagtgtggtac 3780
3805 aaacgcccgtcacagaacaggagggaataataggatggatgacttgacttaagtgtggtac 3864
3781 aggaacagaaaaacttaaggcagtcagcgaagtgcggaagaagagccatcagcgtctcgaat 3840
3865 aggaacagaaaaacttaaggcagtcagcgaagtgcggaagaagagccatcagcgtctcgaat 3924
3841 ctgataacacagcagtggtcctgaggaagagcgtcaaaagaataatttagaaaaatgaag 3900
3925 ctgataacacagcagtggtcctgaggaagagcgtcaaaagaataatttagaaaaatgaag 3984
3901 atgaacagaataatgcgcgaacaaaggggttaaaagagggccgacccacaaacctcttggtg 3960
3985 atgaacagaataatgcgcgaacaaaggggttaaaagagggccgacccacaaacctcttggtg 4044
3961 gagggtacacccaaaaagaagccaaactgaaactcttaaaaaaggaagcaaaaaaaat 4020
4045 gagggtacacccaaaaagaagccaaactgaaactcttaaaaaaggaagcaaaaaaaat 4104

CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
CC located on chromosome 13 at position 13q12-13q. AS3 has a role in
CC inhibiting cell proliferation and use as a marker for the efficient
CC diagnosis and treatment of prostate cancer. The invention includes AS3
CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host
CC cell transfected with the expression vector, and a method for producing
CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC cytosolic activity, and acts to suppress cell proliferation. The AS3
CC gene is useful as a marker for the efficient diagnosis and treatment of
CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC antisense agents for sequence specific modulation of gene expression. The
CC AS3 protein may be used in the treatment of disorders caused by aberrant
CC modification or mutation of a gene encoding an AS3 protein, misregulation
CC of the AS3 gene or aberrant post-translational modification of the AS3
CC protein. This sequence represents the human AS3 cDNA sequence with an
CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC compared with the claimed AS3 cDNA sequence AAA28051.

XX Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 5271; DB 21; Length 5355;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccggagagcccgagtgagcgagtagcgagtcgcaaccgaggggtagaataattt 60
DB |||||||
QY 61 ctgtcatggtcattcaaaactagggaccatgatgaaaaattacatccgctggg 120
DB |||||||
QY 145 ctgtcatggtcattcaaaactagggaccatgatgaaaaattacatccgctggg 204
QY 121 tcaaggaaatcagataaaatctaaagagagagatggtgagacgattaaagatggtg 180
DB |||||||
QY 205 tcaaggaaatcagataaaatctaaagagagagatggtgagacgattaaagatggtg 264
QY 181 tgaaccttttatgatatgaccagagactctgaagagaaaaagggagctttatttaaac 240
DB |||||||
QY 265 tgaaccttttatgatatgaccagagactctgaagagaaaaagggagctttatttaaac 324
QY 241 tagctttacatctgtctcagatctttttctcaagcatcctggtaaagatgtctcgttac 300
DB |||||||
QY 325 tagctttacatctgtctcagatctttttctcaagcatcctggtaaagatgtctcgttac 384
QY 301 tggtagcctgctcgtctcagatctttttctcaagcatcctggtaaagatgtctcgttac 360
DB |||||||
QY 385 tggtagcctgctcgtctcagatctttttctcaagcatcctggtaaagatgtctcgttac 444
QY 361 cccctgataaactaaaggatatttatgtttatacaagacagatggaaggggctagagg 420
DB |||||||
QY 445 cccctgataaactaaaggatatttatgtttatacaagacagatggaaggggctagagg 504
QY 421 atacaagagcccaattcaataggtattttatttacttgagaacattgcttggtca 480
DB |||||||
QY 505 atacaagagcccaattcaataggtattttatttacttgagaacattgcttggtca 564
QY 481 agtcataacatatgtttgagttagaagatgcaatgaaatttcaaccagctataca 540
DB |||||||
QY 565 agtcataacatatgtttgagttagaagatgcaatgaaatttcaaccagctataca 624
QY 541 gaacctattttcagttataaaacatggccacatcagaagttccatgtcacatggtag 600
DB |||||||
QY 625 gaacctattttcagttataaaacatggccacatcagaagttccatgtcacatggtag 684
QY 601 accttagctctctatttttgaaaggtgatacagtgctcaggagcttttggatacgg 660
DB |||||||
QY 685 accttagctctctatttttgaaaggtgatacagtgctcaggagcttttggatacgg 744
QY 661 ttttagtaaatctggtaccctgctcaataagaatttaacaagcaagcatatgattggcaa 720
DB |||||||
QY 745 ttttagtaaatctggtaccctgctcaataagaatttaacaagcaagcatatgattggcaa 804

QY 721 aggtcttactgaagagacagctcaagctatttgagccatattaccacttttttaac 780
DB |||||||
QY 805 aggtcttactgaagagacagctcaagctatttgagccatattaccacttttttaac 864
QY 781 aggtcttactgaagagacagctcaagctatttgagccatattaccacttttttaac 840
DB |||||||
QY 865 aggtcttactgaagagacagctcaagctatttgagccatattaccacttttttaac 924
QY 841 ttttgagagctcaacatattgatagtcattttgctgctctctgttttaccagcttgaa 900
DB |||||||
QY 925 ttttgagagctcaacatattgatagtcattttgctgctctctgttttaccagcttgaa 984
QY 901 ttaaaattaaagagcaatgataatgagagcgctcacagttgttaaaactactggcaaaa 960
DB |||||||
QY 985 ttaaaattaaagagcaatgataatgagagcgctcacagttgttaaaactactggcaaaa 1044
QY 961 tgtttgggccaaggattcagaatttggctctcctcaaaaagccactttggcagtgctact 1020
DB |||||||
QY 1045 tgtttgggccaaggattcagaatttggctctcctcaaaaagccactttggcagtgctact 1104
QY 1021 tgggcaggttttaatgatccatgtaccatccgctgggaatgtgtgaaatttctagcc 1080
DB |||||||
QY 1105 tgggcaggttttaatgatccatgtaccatccgctgggaatgtgtgaaatttctagcc 1164
QY 1081 attgtctcatgaaccatcctctgatttagcaaaagacttaacagagtatcttaaaagtgaagt 1140
DB |||||||
QY 1165 attgtctcatgaaccatcctctgatttagcaaaagacttaacagagtatcttaaaagtgaagt 1224
QY 1141 cacatgaccttgaggagacttagacatgatttttgtgtcaatagttacagctgcta 1200
DB |||||||
QY 1225 cacatgaccttgaggagacttagacatgatttttgtgtcaatagttacagctgcta 1284
QY 1201 aaaaggatttctctggttcaatgatacacttaattttgtgagagagagacattag 1260
DB |||||||
QY 1285 aaaaggatttctctggttcaatgatacacttaattttgtgagagagagacattag 1344
QY 1261 acaaacgatgagagtcacgcaaaagacatgagtgagacttgcacaaatttataagaagt 1320
DB |||||||
QY 1345 acaaacgatgagagtcacgcaaaagacatgagtgagacttgcacaaatttataagaagt 1404
QY 1321 atgctttacagtcagcagctggaaaagatgctgcacaaacagatgagtgatcgaagaaga 1380
DB |||||||
QY 1405 atgctttacagtcagcagctggaaaagatgctgcacaaacagatgagtgatcgaagaaga 1464
QY 1381 aattgtcacatatattatcaaaaatagatttattgtagctgactacttgttgaacggatct 1440
DB |||||||
QY 1465 aattgtcacatatattatcaaaaatagatttattgtagctgactacttgttgaacggatct 1524
QY 1441 ttgctcaatcatggttctccacaaatttagaaactacagacggatgaaatgcttatatt 1500
DB |||||||
QY 1525 ttgctcaatcatggttctccacaaatttagaaactacagacggatgaaatgcttatatt 1584
QY 1501 acttgtagccacactggatttaaatctgttgaaagcattgaaatggtggaatggtc 1560
DB |||||||
QY 1585 acttgtagccacactggatttaaatctgttgaaagcattgaaatggtggaatggtc 1644
QY 1561 aaaatctgctccgacatcaagtaaaagatttggctgactgatttaagcaaaccccaaacag 1620
DB |||||||
QY 1645 aaaatctgctccgacatcaagtaaaagatttggctgactgatttaagcaaaccccaaacag 1704
QY 1621 atgcccagtgccaaggccatattttcaaaagtgtgttattacaagaatttacctgact 1680
DB |||||||
QY 1705 atgcccagtgccaaggccatattttcaaaagtgtgttattacaagaatttacctgact 1764
QY 1681 ctggttaaggctcaggatttcatgaagaaattcacacaggtgtttagaagatgatgagaaa 1740
DB |||||||
QY 1765 ctggttaaggctcaggatttcatgaagaaattcacacaggtgtttagaagatgatgagaaa 1824
QY 1741 taagaagaagcagttagaagactttagtccaactctcctgcaagcagctgaaggtt 1800
DB |||||||
QY 1825 taagaagaagcagttagaagactttagtccaactctcctgcaagcagctgaaggtt 1884
QY 1801 gtgtgcgtgaaataactaagaagttgggcaaccccccaaacagcctacaaatctcttcctgg 1860

DE Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4.

KW Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW Chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW diagnosis; treatment; cytostatic; human; ss.

OS Homo sapiens.

Key Location/Qualifiers

FT exon 85..130
FT /tag= a
FT /number= 1
FT 131..257
FT /tag= b
FT /number= 2
FT 150..4325
FT /tag= c
FT /product= "AS3"
FT /note= "Androgen shutoff gene 3 protein"
FT exon 258..461
FT /tag= d
FT /number= 3
FT exon 462..548
FT /tag= e
FT /number= 4
FT 549..646
FT /tag= f
FT /number= 5
FT exon 647..773
FT /tag= e
FT /number= 6
FT 774..938
FT /tag= f
FT /number= 7
FT exon 939..995
FT /tag= g
FT /number= 8
FT 996..1111
FT /tag= h
FT /number= 9
FT 1112..1206
FT /tag= i
FT /number= 10
FT exon 1207..1352
FT /tag= j
FT /number= 11
FT exon 1353..1504
FT /tag= k
FT /number= 12
FT exon 1505..1618
FT /tag= l
FT /number= 13
FT exon 1619..1700
FT /tag= m
FT /number= 14
FT exon 1701..1749
FT /tag= n
FT /number= 15
FT exon 1750..1889
FT /tag= o
FT /number= 16
FT exon 1890..2005
FT /tag= p
FT /number= 17
FT exon 2006..2111
FT /tag= q
FT /number= 18
FT 2112..2272
FT /tag= r
FT /number= 19
FT exon 2273..2396
FT /tag= s
FT /number= 20

FT exon 2397..2555
FT /tag= t
FT /number= 21
FT 2556..2624
FT /tag= u
FT /number= 22
FT 2625..2761
FT /tag= v
FT /number= 23
FT 2672..2885
FT /tag= w
FT /number= 24
FT 2886..3090
FT /tag= x
FT /number= 25
FT 3091..3205
FT /tag= y
FT /number= 26
FT 3206..3338
FT /tag= z
FT /number= 27
FT 3339..3458
FT /tag= aa
FT /number= 28
FT 3459..3521
FT /tag= ab
FT /number= 29
FT 3522..3667
FT /tag= ac
FT /number= 30
FT 3668..3773
FT /tag= ad
FT /number= 31
FT 3774..4213
FT /tag= ae
FT /number= 32
FT 4214..4438
FT /tag= af
FT /number= 33
FT misc_signal 4314..4408
FT /tag= ag
FT /note= "Destabilizing signal"
FT exon 4439..5337
FT /tag= ah
FT /number= 34
FT polyA_signal 5312..5317
FT /tag= ai
FT misc_feature 5333..5337
FT /tag= aj
FT /note= "Poly adenylation cleavage site"
FT exon
FT WO2000050454-A1.
FT 31-AUG-2000.
FT 24-FEB-2000; 2000WO-US04732.
FT XX
FT 24-FEB-1999; 99US-0121461.
FT (TUFT) TUFTS COLLEGE.
FT Soto AM, Sonnenschein C, Geck P, Szelel J;
FT WPI; 2000-565451/52.
FT P-PSDB; AAY94702.
FT New human androgen-induced tumor suppressor cDNA sequence termed
FT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
FT diagnosis and treatment of prostate cancer -
FT Example 4; Fig 6; 152pp; English.
FT PS
FT This invention relates to a human androgen-induced tumour suppressor
FT CC


```

FT FT /*tag= z
FT FT /number= 27
FT FT exon 3255..3374
FT FT /*tag= aa
FT FT /number= 28
FT FT exon 3375..3437
FT FT /*tag= ab
FT FT /number= 29
FT FT exon 3438..3583
FT FT /*tag= ac
FT FT /number= 30
FT FT exon 3584..3689
FT FT /*tag= ad
FT FT /number= 31
FT FT exon 3690..4129
FT FT /*tag= ae
FT FT /number= 32
FT FT exon 4130..4354
FT FT /*tag= af
FT FT /number= 33
FT FT misc_signal 4320..4324
FT FT /*tag= ag
FT FT /note= "Destabilizing signal"
FT FT exon 4355..5253
FT FT /*tag= ah
FT FT /number= 34
FT FT polyA_signal 5228..5233
FT FT /*tag= ai
FT FT misc_feature 5249..5253
FT FT /*tag= aj
FT FT /note= "Poly adenylation cleavage site"
XX XX
XX XX WO200050454-A1.
XX XX
XX XX 31-AUG-2000.
XX XX
XX XX 24-FEB-2000; 2000WO-US04732.
XX XX
XX XX 24-FEB-1999; 99US-0121461.
XX XX
XX XX (TUFT ) TUFTS COLLEGE.
XX XX
XX XX Soto AM, Sonnenschein C, Geck P, Szelei J;
XX XX
XX XX WPI; 2000-565451/52.
XX XX P-PSDB; AAY94702.
XX XX
XX XX New human androgen-induced tumor suppressor cDNA sequence termed
XX XX 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
XX XX diagnosis and treatment of prostate cancer -
XX XX
XX XX Claim 1; Fig 1; 152pp; English.
XX XX
XX XX This invention relates to a human androgen-induced tumour suppressor cDNA
XX XX sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
XX XX on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
XX XX proliferation and use as a marker for the efficient diagnosis and
XX XX treatment of prostate cancer. The invention includes AS3 cDNA and protein
XX XX sequences, a vector comprising the cDNA sequence, a host cell transfected
XX XX with the expression vector, and a method for producing an AS3 polypeptide
XX XX comprising culturing the transfected cells. AS3 has cytostatic activity,
XX XX and acts to suppress cell proliferation. The AS3 gene is useful as a
XX XX marker for the efficient diagnosis and treatment of prostate cancer. The
XX XX AS3 nucleic acid molecule can be used as a source of antisense agents for
XX XX sequence specific modulation of gene expression. The AS3 protein may be
XX XX used in the treatment of disorders caused by aberrant modification or
XX XX mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
XX XX or aberrant post-translational modification of the AS3 protein. This
XX XX sequence represents the human AS3 cDNA sequence.
XX XX
XX XX Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;

```

```

Query Match 100.0%; Score 5271; DB 21; Length 5271;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cggagagcccgagtgagcgagtagcgagtcggcaaccggagggtagaataattt 60
Db 1 cggagagcccgagtgagcgagtagcgagtcggcaaccggagggtagaataattt 60

QY 61 ctgtcatgctcattcaaaagactaggaccatgatgaaataattacataatccgctggg 120
Db 61 ctgtcatgctcattcaaaagactaggaccatgatgaaataattacataatccgctggg 120

QY 121 tcaaggaatatcagataaaatatctaaagagagatggtgagacgattaaagatggtg 180
Db 121 tcaaggaatatcagataaaatatctaaagagagatggtgagacgattaaagatggtg 180

QY 181 tgaaccttttatggatatgaccagagactctgaagaagaaaaggagctttatttaacc 240
Db 181 tgaaccttttatggatatgaccagagactctgaagaagaaaaggagctttatttaacc 240

QY 241 tagcttttaaatctgtcttcagatttttttcaagaatcctcgttaagatgttcgcttac 300
Db 241 tagcttttaaatctgtcttcagatttttttcaagaatcctcgttaagatgttcgcttac 300

QY 301 tggtagcctgctgctgtgatatatttcaggatttatgctcctgaagctcctcacat 360
Db 301 tggtagcctgctgctgtgatatatttcaggatttatgctcctgaagctcctcacat 360

QY 361 cccctgataaactaaaggatatatttatgtttatacaaacagacagttgaagggttagagg 420
Db 361 cccctgataaactaaaggatatatttatgtttatacaaacagacagttgaagggttagagg 420

QY 421 atacaagagcccaaatccaataggtatttttatttacttgagaacatgtctgggtca 480
Db 421 atacaagagcccaaatccaataggtatttttatttacttgagaacatgtctgggtca 480

QY 481 agtcataacatatgctttgagtagaagatagcaatgaaattttcaccagctataca 540
Db 481 agtcataacatatgctttgagtagaagatagcaatgaaattttcaccagctataca 540

QY 541 gaaccttattttcagttataaacaatggccacaatgcagaagtcctatgcacatggtag 600
Db 541 gaaccttattttcagttataaacaatggccacaatgcagaagtcctatgcacatggtag 600

QY 601 acctatgagctctattatttgaaggtagacagtgctcgaaggagcttttgatcagg 660
Db 601 acctatgagctctattatttgaaggtagacagtgctcgaaggagcttttgatcagg 660

QY 661 ttttagtaaatctggtaccctgctcataagaatttaacaagcaagcatatgattggcaa 720
Db 661 ttttagtaaatctggtaccctgctcataagaatttaacaagcaagcatatgattggcaa 720

QY 721 aggtctgagctctcaaatatgatgacatctatgcgtctctgttttaccocagcttgaat 900
Db 721 aggtctgagctctcaaatatgatgacatctatgcgtctctgttttaccocagcttgaat 900

QY 901 ttaaataaagagcaatgataatgagagcgccctacagttgtttaaactactgcgcaaaa 960
Db 901 ttaaataaagagcaatgataatgagagcgccctacagttgtttaaactactgcgcaaaa 960

QY 961 tgtttggggcaaaagattcagaattggtctctcaaaacagccactttggcagtgctact 1020
Db 961 tgtttggggcaaaagattcagaattggtctctcaaaacagccactttggcagtgctact 1020

QY 1021 tgggcagggttttaagtatccatgataccaatcccgctggaatgtgcaatttgcctagcc 1080

```


C 961	24	0.5	1671	22	AAF33038	Human secreted pro	FT	exon	174...377
C 962	24	0.5	1673	21	AAZ65064	Membrane-bound pro	FT		/tag= d
C 963	24	0.5	1674	21	AAZ58633	Human PRO336 prote	FT		/number= 3
C 964	24	0.5	1674	22	AAF44210	Human PRO336 (UNQ5	FT	exon	378...464
C 965	24	0.5	1678	21	AAF16140	Human prostate can	FT		/tag= e
C 966	24	0.5	1681	22	AAF33135	Human secreted pro	FT		/number= 4
C 967	24	0.5	1688	15	AAQ64613	Human mdr-1 promot	FT	exon	465...562
C 968	24	0.5	1693	20	AAV69206	Canine ICAM-R poly	FT		/tag= f
C 969	24	0.5	1693	21	AAZ24362	Canine ICAM-R DNA	FT		/number= 5
C 970	24	0.5	1700	21	AAZ52505	Human secreted pro	FT	exon	563...689
C 971	24	0.5	1704	22	AAC91180	ADAM gene #4. Hom	FT		/tag= e
C 972	24	0.5	1705	16	AAQ79621	Schizosaccharomyce	FT		/number= 6
C 973	24	0.5	1705	22	AAF44264	Human PRO3301 nucl	FT	exon	690...854
C 974	24	0.5	1708	22	AAF26570	DNA encoding human	FT		/tag= f
C 975	24	0.5	1720	20	AAV84533	Human secreted pro	FT		/number= 7
C 976	24	0.5	1724	21	AAA16694	Human secreted pro	FT	exon	855...911
C 977	24	0.5	1733	19	AAV20607	Rat kynurenine-3-h	FT		/tag= g
C 978	24	0.5	1733	20	AAV81394	Human tumour antig	FT		/number= 8
C 979	24	0.5	1733	21	AAZ33178	Corn delta-12 deha	FT	exon	912...1027
C 980	24	0.5	1738	21	AAC99089	Human pancreatic c	FT		/tag= h
C 981	24	0.5	1751	21	AAC68122	Human secreted pro	FT		/number= 9
C 982	24	0.5	1752	21	AAZ33348	Human secreted pro	FT	exon	1028...1122
C 983	24	0.5	1756	21	AAF18203	Lung cancer associ	FT		/tag= i
C 984	24	0.5	1758	22	AAF32703	Human secreted pro	FT		/number= 10
C 985	24	0.5	1762	20	AAZ58801	Human leukaemia-as	FT	exon	1123...1268
C 986	24	0.5	1767	20	AAZ51749	DNA encoding a hum	FT		/tag= j
C 987	24	0.5	1770	14	AAQ37684	P.falciparum antig	FT		/number= 11
C 988	24	0.5	1770	14	AAQ37124	P.falciparum antig	FT	exon	1269...1420
C 989	24	0.5	1772	21	AAF21809	Human breast and o	FT		/tag= k
C 990	24	0.5	1776	21	AAA49177	CDNA encoding huma	FT		/number= 12
C 991	24	0.5	1781	14	AAQ46991	CDNA clone encodin	FT	exon	1421...1534
C 992	24	0.5	1781	19	AAV56348	Human ICAM-R CDNA.	FT		/tag= l
C 993	24	0.5	1781	19	AAV54826	CDNA encoding huma	FT		/number= 13
C 994	24	0.5	1781	19	AAV11657	Human ICAM-R CDNA.	FT	exon	1535...1616
C 995	24	0.5	1781	19	AAV19328	Human ICAM-4 CDNA.	FT		/tag= m
C 996	24	0.5	1781	20	AAZ21855	Human ICAM-R codin	FT		/number= 14
C 997	24	0.5	1781	20	AAV69125	Intercellular adhe	FT	exon	1617...1665
C 998	24	0.5	1781	21	AAA97090	Human CDNA encodin	FT		/tag= n
C 999	24	0.5	1781	21	AAA08236	Human ICAM-R prote	FT	exon	1666...1805
C1000	24	0.5	1781	21	AAZ24262	Human ICAM-R CDNA.	FT		/tag= o
							FT		/number= 16
							FT	exon	1806...1921
							FT		/tag= p
							FT		/number= 17
							FT	exon	1922...2027
							FT		/tag= q
							FT		/number= 18
							FT	exon	2028...2188
							FT		/tag= r
							FT		/number= 19
							FT	exon	2189...2312
							FT		/tag= s
							FT		/number= 20
							FT	exon	2313...2471
							FT		/tag= t
							FT		/number= 21
							FT	exon	2472...2540
							FT		/tag= u
							FT		/number= 22
							FT	exon	2541...2677
							FT		/tag= v
							FT		/number= 23
							FT	exon	2678...2801
							FT		/tag= w
							FT		/number= 24
							FT	exon	2802...3006
							FT		/tag= x
							FT		/number= 25
							FT	exon	3007...3121
							FT		/tag= y
							FT		/number= 26
							FT	exon	3122...3254

ALIGNMENTS

RESULT 1
AAA28051
ID AAA28051 standard; cDNA; 5271 BP.
XX
AC AAA28051;
XX
DT 01-DRC-2000 (first entry)
XX Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX
DE Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
XX chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW diagnosis; treatment; cytostatic; human; ss.
KW
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT exon 1..46
FT /tag= a
FT /number= 1
FT 47..173
FT /tag= b
FT /number= 2
FT 66..4241
FT /tag= c
FT /product= "AS3"
FT /note= "Androgen shutoff gene 3 protein, the CDS is
FT specifically claimed as SEQ ID #3".
FT

c 815	24	0.5	1257	13	AAQ24292	(1-3,1-4)-beta-glu	c 888	24	0.5	1462	20	AA241241	Human normal ovari
c 816	24	0.5	1260	11	AA250084	Impatiens balsamia	c 889	24	0.5	1469	14	AAQ40323	Sequence Of the CD
c 817	24	0.5	1270	16	AAQ87101	The Tpa gene, Syn	c 890	24	0.5	1475	21	AAC59122	Human secreted pro
c 818	24	0.5	1270	18	AA763685	Full-length Tpa-1	c 891	24	0.5	1478	21	AA241405	Human normal pancr
c 819	24	0.5	1270	20	AA401310	Human normal ovari	c 892	24	0.5	1481	21	AAC69406	Human secreted pro
c 820	24	0.5	1270	20	AA401390	Full length anioni	c 893	24	0.5	1486	21	AA297081	Human secreted pro
c 821	24	0.5	1276	17	AA742071	Human nuclear prol	c 894	24	0.5	1493	13	AAQ28411	Glia activating fa
c 822	24	0.5	1276	21	AA797977	Human secreted pro	c 895	24	0.5	1497	22	AAE27793	Human TANGO 295 co
c 823	24	0.5	1278	21	AA21698	Human breast and o	c 896	24	0.5	1500	21	AAC53986	Arabidopsis thalia
c 824	24	0.5	1278	22	AA277543	Caenorhabditis ele	c 897	24	0.5	1508	21	AA37025	Human PRO1773 (UNQ
c 825	24	0.5	1282	21	AAC93379	Human secreted pro	c 898	24	0.5	1508	22	AA292067	Human PRO1773 CDNA
c 826	24	0.5	1284	20	AA252938	Human prostate tum	c 899	24	0.5	1508	22	AA254207	DNA encoding prote
c 827	24	0.5	1284	20	AA221094	Human secreted pro	c 900	24	0.5	1509	19	AAV59679	Human secreted pro
c 828	24	0.5	1293	21	AA293129	Human secreted pro	c 901	24	0.5	1516	21	AAC69497	Human secreted pro
c 829	24	0.5	1295	21	AA261171	Regulatory sequence	c 902	24	0.5	1516	21	AA594440	Nucleotide sequenc
c 830	24	0.5	1297	13	AAQ21645	3' coding sequence	c 903	24	0.5	1520	8	AA270223	Plasmid pBRV2 inse
c 831	24	0.5	1299	21	AA116866	Human secreted pro	c 904	24	0.5	1520	8	AA271320	Sequence encoding
c 832	24	0.5	1300	18	AA764848	Meripilus giganteu	c 905	24	0.5	1521	7	AA260937	Plasmid pBRV2 inse
c 833	24	0.5	1300	21	AAC81109	Human secreted pro	c 906	24	0.5	1522	19	AAV46309	Human secreted pro
c 834	24	0.5	1300	21	AAC81109	Human secreted pro	c 907	24	0.5	1522	19	AAV46309	Human secreted pro
c 835	24	0.5	1302	20	AA290446	Human secreted pro	c 908	24	0.5	1522	21	AA37027	Nucleotide sequenc
c 836	24	0.5	1303	20	AA242221	Human normal bladd	c 909	24	0.5	1524	22	AA254212	Human PRO1478 (UNQ
c 837	24	0.5	1305	19	AAV26270	Trichomonas vagina	c 910	24	0.5	1524	22	AAQ31200	DNA encoding prote
c 838	24	0.5	1305	19	AAV26272	Trichomonas vagina	c 911	24	0.5	1525	13	AAQ31200	Myo-inositol O-met
c 839	24	0.5	1315	20	AA241312	Human normal ovari	c 912	24	0.5	1525	18	AA295213	cDNA encoding the
c 840	24	0.5	1320	16	AA229190	Glycoprotein 39 3'	c 913	24	0.5	1525	19	AAV22704	Mouse recombinase
c 841	24	0.5	1320	19	AAV20948	Coffee-fruit speci	c 914	24	0.5	1525	21	AA27911	CDNA encoding IL-1
c 842	24	0.5	1326	20	AA334867	CDNA encoding Brus	c 915	24	0.5	1525	21	AAQ08554	Murine REC2 serine
c 843	24	0.5	1328	19	AAV58754	Human secreted pro	c 916	24	0.5	1525	22	AAQ02334	Murine interleukin
c 844	24	0.5	1334	21	AAV52490	Human MN gene intr	c 917	24	0.5	1525	22	AAC81415	Mouse IL-13 recept
c 845	24	0.5	1334	21	AA116571	Human MN intron 6	c 918	24	0.5	1529	21	AAC59838	Human secreted pro
c 846	24	0.5	1338	20	AA297959	Human secreted pro	c 919	24	0.5	1530	8	AA270222	Plasmid pBRG4 inse
c 847	24	0.5	1338	21	AA276985	Human ORFX ORF2540	c 920	24	0.5	1531	21	AA297369	Plasmid pBRG4 inse
c 848	24	0.5	1341	20	AA257178	Human KCNQ3 intron	c 921	24	0.5	1533	20	AAV84490	Human colorectal c
c 849	24	0.5	1342	21	AA26424	Human secreted pro	c 922	24	0.5	1533	22	AA232702	Human secreted pro
c 850	24	0.5	1348	15	AAQ78278	T. niveum GAPDH	c 923	24	0.5	1534	8	AA271319	Sequence encoding
c 851	24	0.5	1348	21	AA550401	Arabidopsis thalia	c 924	24	0.5	1535	21	AA21650	Human breast and o
c 852	24	0.5	1353	22	AA291863	Human secreted pro	c 925	24	0.5	1537	21	AA279967	Human secreted pro
c 853	24	0.5	1355	9	AA280457	Sequence of cDNA e	c 926	24	0.5	1537	21	AA279967	Human secreted pro
c 854	24	0.5	1355	20	AA207164	Human lung tumour	c 927	24	0.5	1538	20	AAV84593	Human secreted pro
c 855	24	0.5	1355	21	AA279076	Human lung tumour-	c 928	24	0.5	1538	21	AAQ08416	Human tumour suppr
c 856	24	0.5	1356	20	AA297957	Human secreted pro	c 929	24	0.5	1539	20	AA233317	Human PDSP-1 DNA
c 857	24	0.5	1358	22	AAC85033	Soybean HES1 homol	c 930	24	0.5	1547	20	AA333448	Oryza sativa L. pi
c 858	24	0.5	1359	21	AAC66430	Human secreted pro	c 931	24	0.5	1547	21	AAC59424	Human secreted pro
c 859	24	0.5	1360	21	AA299029	Human TGC839 nucle	c 932	24	0.5	1549	21	AAC77188	Human ORFX ORF2743
c 860	24	0.5	1371	19	AAV33755	Mouse neutrophil e	c 933	24	0.5	1551	19	AAV03028	Globodera rostochl
c 861	24	0.5	1374	21	AA339407	Corn SR2 homology	c 934	24	0.5	1554	20	AA216675	Xenopus WA545 prot
c 862	24	0.5	1376	15	AAQ56747	Ryegrass Lol pv al	c 935	24	0.5	1558	18	AA222942	Human cathepsin y
c 863	24	0.5	1377	21	AAQ55190	Human secreted pro	c 936	24	0.5	1559	21	AAC47168	Mouse sphingosine
c 864	24	0.5	1378	19	AAV59706	Human secreted pro	c 937	24	0.5	1571	21	AAC36366	Arabidopsis thalia
c 865	24	0.5	1381	21	AA448572	Soybean protein ph	c 938	24	0.5	1573	19	AAV43607	Human secreted pro
c 866	24	0.5	1386	19	AAV11855	Homo sapiens Tub I	c 939	24	0.5	1573	21	AA54286	Human mesenchymal
c 867	24	0.5	1386	21	AAC77185	Human ORFX ORF2740	c 940	24	0.5	1574	21	AA265325	Human secreted pro
c 868	24	0.5	1387	22	AAQ02201	Chondrus crispus s	c 941	24	0.5	1580	21	AA261288	Human secreted pro
c 869	24	0.5	1391	17	AA270329	Human calcium sign	c 942	24	0.5	1589	21	AA295198	Human UGT2B4 exon
c 870	24	0.5	1396	21	AA298027	Human secreted pro	c 943	24	0.5	1592	21	AAC77050	Human ORFX ORF2605
c 871	24	0.5	1404	20	AAV69207	DNA sequence of ca	c 944	24	0.5	1607	18	AAV85878	Malassezia fungus
c 872	24	0.5	1404	21	AA224363	Canine ICAM-R DNA	c 945	24	0.5	1610	20	AA252902	Human prostate tum
c 873	24	0.5	1405	20	AA280908	Human cDNA clone H	c 946	24	0.5	1614	21	AA221791	Human breast and o
c 874	24	0.5	1407	7	AA260374	Sequence encoding	c 947	24	0.5	1616	21	AA252506	Human secreted pro
c 875	24	0.5	1407	14	AAQ52515	Human phospholipas	c 948	24	0.5	1623	19	AAV07059	Mouse mitogen acti
c 876	24	0.5	1407	15	AAQ71647	Human phospholipas	c 949	24	0.5	1624	20	AAV09669	Human cathepsin K
c 877	24	0.5	1407	18	AA289675	Human breast and o	c 950	24	0.5	1624	20	AAV84570	Human secreted pro
c 878	24	0.5	1408	21	AA221863	Human colon cancer	c 951	24	0.5	1631	18	AAV72795	Beta-1 integrin mo
c 879	24	0.5	1410	21	AAC98057	Human ovarian tumo	c 952	24	0.5	1637	21	AAC77959	Human cancer assoc
c 880	24	0.5	1427	20	AA277561	Maize myo-inositol	c 953	24	0.5	1637	22	AA275093	Human colon associ
c 881	24	0.5	1428	20	AA244403	Thiosulphate sulph	c 954	24	0.5	1642	19	AAV40507	Human secreted pro
c 882	24	0.5	1438	21	AA293098	Human cancer assoc	c 955	24	0.5	1643	20	AAV07062	Mouse mitogen acti
c 883	24	0.5	1446	21	AAC77994	Plasmodium falci	c 956	24	0.5	1647	21	AA213841	Aspergillus fumiga
c 884	24	0.5	1452	21	AA270123	Human ORFX ORF1592	c 957	24	0.5	1651	21	AA215694	Cancer suppressor
c 885	24	0.5	1457	21	AAC76037	CDNA encoding huma	c 958	24	0.5	1651	22	AA297914	Human secreted pro
c 886	24	0.5	1457	21	AA234553	Human ORFX ORF2978	c 959	24	0.5	1670	19	AAV46310	Human secreted pro
c 887	24	0.5	1459	21	AAC77423		c 960	24	0.5	1670	20	AAV84836	CDNA molecule prod

669	24	0.5	867	18	AA66791	Partial seq. obtai	742	24	0.5	1020	21	AAZ49558	Partial DNA sequen
670	24	0.5	870	17	AA32400	Flea allergen cDNA	743	24	0.5	1021	21	AAZ97100	Human secreted pro
671	24	0.5	874	20	AAZ22219	Human secreted pro	c 744	24	0.5	1024	21	AAZ57777	Human ORFX ORF1332
672	24	0.5	877	19	AAV04635	Human cytoplasmic	745	24	0.5	1026	21	AAZ97206	Human prostate can
673	24	0.5	878	21	AAZ94097	Haematopoietic ste	746	24	0.5	1026	21	AAZ77906	Human cancer assoc
674	24	0.5	884	20	AAZ41356	Human normal uteru	c 747	24	0.5	1027	20	AAZ84690	CBCALD05 gene. Ho
675	24	0.5	885	21	AAZ78135	Human cancer assoc	c 748	24	0.5	1027	21	AAZ98781	Human pancreatic c
676	24	0.5	888	21	AAZ52552	Human secreted pro	c 749	24	0.5	1029	21	AAZ15949	Human prostate can
677	24	0.5	891	20	AAZ06233	Human secreted pro	c 750	24	0.5	1031	20	AAZ24419	Human bladder tumo
678	24	0.5	893	20	AAZ08308	Human lung tumour-	c 751	24	0.5	1033	17	AAZ36363	High mobility grou
679	24	0.5	893	21	AAZ79160	Human lung tumour-	c 752	24	0.5	1033	21	AAZ89462	Corn leud subunit
680	24	0.5	893	22	AAZ26575	DNA encoding human	c 753	24	0.5	1038	20	AAZ27247	Human secreted pro
681	24	0.5	894	21	AAZ64640	Partial sequence M	c 754	24	0.5	1040	16	AAZ92530	P. communis (pear)
682	24	0.5	896	20	AAZ41976	Human myometrium t	c 755	24	0.5	1040	20	AAZ58275	Zea mays SCLBr pro
683	24	0.5	899	20	AAZ15217	Human gene express	c 756	24	0.5	1040	21	AAZ96932	Human secreted pro
684	24	0.5	900	21	AAZ79020	Human secreted pro	c 757	24	0.5	1050	22	AAZ86483	Maize Mac2 coding
685	24	0.5	905	19	AAZ69627	Human secreted pro	c 758	24	0.5	1054	21	AAZ95529	Human secreted pro
686	24	0.5	907	19	AAZ59695	Human secreted pro	c 759	24	0.5	1060	20	AAZ84447	Mouse brain CNG-4
687	24	0.5	907	20	AAZ35739	DNA encoding a pr	c 760	24	0.5	1064	16	AAZ87093	Tpa-1 phase inser
688	24	0.5	911	8	AAZ11002	Sequence encoding	c 761	24	0.5	1064	18	AAZ63677	Peroxidase gene, T
689	24	0.5	919	20	AAZ07555	Maize glutathione-	c 762	24	0.5	1064	20	AAZ03962	Anionic tobacco pe
690	24	0.5	919	21	AAZ76260	Maize glutathione-	c 763	24	0.5	1069	20	AAZ84474	Human secreted pro
691	24	0.5	919	21	AAZ94795	Maize class II glu	c 764	24	0.5	1071	20	AAZ20414	Human secreted pro
692	24	0.5	920	21	AAZ59357	Human secreted pro	c 765	24	0.5	1074	22	AAZ23934	Human secreted pro
693	24	0.5	925	21	AAZ52534	Human secreted pro	c 766	24	0.5	1082	20	AAZ41967	Human myometrium t
694	24	0.5	926	21	AAZ59967	Human secreted pro	c 767	24	0.5	1083	20	AAZ10659	DNA encoding a hu
695	24	0.5	926	22	AAZ30548	Soybean FLF MADS-b	c 768	24	0.5	1097	22	AAZ02653	Tomato macrocalyx
696	24	0.5	928	20	AAZ20207	Soybean raffinose	c 769	24	0.5	1106	21	AAZ29167	Rice 4-hydroxyphen
697	24	0.5	928	20	AAZ61395	DNA encoding a hum	c 770	24	0.5	1109	22	AAZ00288	M. cinnamomea xylo
698	24	0.5	928	21	AAZ74443	Human secreted pro	c 771	24	0.5	1111	20	AAZ52949	Human prostate tum
699	24	0.5	928	21	AAZ69576	Human secreted pro	c 772	24	0.5	1116	21	AAZ46246	Amaranthus viridis
700	24	0.5	943	21	AAZ79888	Human secreted pro	c 773	24	0.5	1117	17	AAZ37080	AR192 fragment enc
701	24	0.5	943	21	AAZ98022	Human secreted pro	c 774	24	0.5	1118	21	AAZ52547	Human secreted pro
702	24	0.5	944	21	AAZ97113	Human secreted pro	c 775	24	0.5	1122	21	AAZ59020	Human secreted pro
703	24	0.5	946	21	AAZ26294	Human secreted pro	c 776	24	0.5	1126	21	AAZ59409	Human secreted pro
704	24	0.5	947	22	AAZ59084	Mouse MLP nucleoti	c 777	24	0.5	1128	18	AAZ86027	Meripilus giganteu
705	24	0.5	952	21	AAZ78431	Human secreted pro	c 778	24	0.5	1134	21	AAZ74213	DNA encoding eviro
706	24	0.5	954	20	AAZ37477	Human secreted pro	c 779	24	0.5	1138	22	AAZ65061	Membrane-bound pro
707	24	0.5	959	21	AAZ26370	Human secreted pro	c 780	24	0.5	1140	22	AAZ62578	Human secreted pro
708	24	0.5	960	21	AAZ61884	DNA encoding a hu	c 781	24	0.5	1141	21	AAZ59437	Human secreted pro
709	24	0.5	963	21	AAZ22332	Human secreted pro	c 782	24	0.5	1142	15	AAZ67320	Fisher rat Platele
710	24	0.5	963	22	AAZ44895	Human breast cance	c 783	24	0.5	1148	22	AAZ74213	DNA encoding eviro
711	24	0.5	964	20	AAZ06773	Marmoset sperm-spe	c 784	24	0.5	1164	21	AAZ65061	Membrane-bound pro
712	24	0.5	968	21	AAZ98075	Human colon cancer	c 785	24	0.5	1164	22	AAZ44207	Human PRO511 (UNQ5
713	24	0.5	970	20	AAZ52871	Human prostate tum	c 786	24	0.5	1167	22	AAZ64199	Human secreted pro
714	24	0.5	970	21	AAZ98172	Human colon cancer	c 787	24	0.5	1174	20	AAZ21516	Progression suppl
715	24	0.5	972	19	AAZ11784	H. contortus pHC-2	c 788	24	0.5	1174	20	AAZ61432	DNA encoding a hum
716	24	0.5	973	20	AAZ21192	Zea mays pathogene	c 789	24	0.5	1175	21	AAZ73905	Human secreted pro
717	24	0.5	975	19	AAZ61499	Bak binding protei	c 790	24	0.5	1178	21	AAZ90469	Human uncoupling p
718	24	0.5	976	18	AAZ63288	Partial human tryp	c 791	24	0.5	1179	19	AAZ59761	Human secreted pro
719	24	0.5	976	18	AAZ63288	Partial human tryp	c 792	24	0.5	1180	20	AAZ84347	Human Apo-2bcr cDN
720	24	0.5	977	21	AAZ52526	Human secreted pro	c 793	24	0.5	1187	21	AAZ75882	Human ORFX ORF1437
721	24	0.5	982	21	AAZ59197	Human secreted pro	c 794	24	0.5	1190	21	AAZ78384	Human secreted pro
722	24	0.5	983	19	AAZ17089	Homo sapiens tubul	c 795	24	0.5	1195	21	AAZ94943	Human carbohydrate
723	24	0.5	984	21	AAZ15737	Human prostate can	c 796	24	0.5	1200	11	AAZ05728	Gene encoding glut
724	24	0.5	985	20	AAZ27250	Human secreted pro	c 797	24	0.5	1200	21	AAZ15723	Human prostate can
725	24	0.5	988	20	AAZ61428	DNA encoding a hum	c 798	24	0.5	1200	21	AAZ54293	Human mesenchymal
726	24	0.5	989	20	AAZ24820	Human secreted pro	c 799	24	0.5	1212	20	AAZ76499	Human WISP-3 prote
727	24	0.5	995	20	AAZ59354	Human secreted pro	c 800	24	0.5	1212	20	AAZ76500	Human WISP-3 prote
728	24	0.5	995	21	AAZ69107	Human secreted pro	c 801	24	0.5	1213	21	AAZ77817	Human cancer assoc
729	24	0.5	996	16	AAZ89606	Human membrane pre	c 802	24	0.5	1215	20	AAZ00620	Human secreted pro
730	24	0.5	997	20	AAZ41269	Human normal ovari	c 803	24	0.5	1216	21	AAZ98087	Human colon cancer
731	24	0.5	1000	21	AAZ77786	Human cancer assoc	c 804	24	0.5	1219	21	AAZ23434	cDNA encoding huma
732	24	0.5	1001	21	AAZ57481	Arachidonic acid m	c 805	24	0.5	1219	22	AAZ64202	Human secreted pro
733	24	0.5	1001	21	AAZ57995	Arachidonic acid m	c 806	24	0.5	1232	21	AAZ73867	Human secreted pro
734	24	0.5	1001	21	AAZ57996	Arachidonic acid m	c 807	24	0.5	1235	21	AAZ52509	Human secreted pro
735	24	0.5	1003	22	AAZ91893	Human secreted pro	c 808	24	0.5	1236	9	AAZ80229	Bioadhesive precu
736	24	0.5	1007	21	AAZ95508	Human secreted pro	c 809	24	0.5	1236	9	AAZ80229	Sequence for a gen
737	24	0.5	1013	12	AAZ12431	Fragment D of urat	c 810	24	0.5	1237	21	AAZ77309	Human ORFX ORF2864
738	24	0.5	1016	20	AAZ77920	Human secreted pro	c 811	24	0.5	1240	21	AAZ98130	Human colon cancer
739	24	0.5	1018	20	AAZ77486	Human ovarian tumo	c 812	24	0.5	1247	21	AAZ59186	Human secreted pro
740	24	0.5	1020	20	AAZ52865	Human prostate tum	c 813	24	0.5	1253	16	AAZ82973	Cluster-A cDNA enc
741	24	0.5	1020	20	AAZ81240	Human oligophrenin	c 814	24	0.5	1254	21	AAZ95463	Human secreted pro

523	24	0.5	566	21	AAZ80382	Human colon cancer
524	24	0.5	567	20	AAZ51738	DNA encoding a hum
525	24	0.5	569	21	AAZ97052	Human secreted pro
526	24	0.5	571	20	AAZ61364	DNA encoding a hum
527	24	0.5	572	19	AAV44301	Human secreted pro
528	24	0.5	575	20	AAV84619	Human secreted pro
529	24	0.5	578	21	AAZ76861	Human ORFX ORF2416
530	24	0.5	578	21	AAZ80673	Human colon cancer
531	24	0.5	580	17	AAZ29031	Endoglucanase (Fam
532	24	0.5	581	21	AAZ15634	Human prostate can
533	24	0.5	581	21	AAZ16007	Human colon cancer
534	24	0.5	587	21	AAZ80546	Human secreted pro
535	24	0.5	588	20	AAV84464	Human secreted pro
536	24	0.5	591	21	AAZ16401	Human colon cancer
537	24	0.5	592	22	AAZ74202	DNA encoding eviro
538	24	0.5	595	20	AAZ41303	Human normal ovari
539	24	0.5	599	21	AAZ80063	Human colon cancer
540	24	0.5	602	20	AAZ41243	Human normal ovari
541	24	0.5	603	20	AAZ20419	Human secreted pro
542	24	0.5	603	18	AAZ91169	Secreted cytokine
543	24	0.5	603	21	AAZ75983	Human ORFX ORF1538
544	24	0.5	604	21	AAZ80307	Human colon cancer
545	24	0.5	606	20	AAZ24555	Human lung tumor a
546	24	0.5	606	21	AAZ65794	Human lung cancer-
547	24	0.5	607	21	AAZ16283	Human colon cancer
548	24	0.5	608	20	AAZ36688	Mouse OHP106 proti
549	24	0.5	608	21	AAZ16178	Human colon cancer
550	24	0.5	608	21	AAZ16518	Human colon cancer
551	24	0.5	611	21	AAZ16173	Human colon cancer
552	24	0.5	612	21	AAZ16275	Human colon cancer
553	24	0.5	620	20	AAZ06238	Human secreted pro
554	24	0.5	622	21	AAZ16179	Human colon cancer
555	24	0.5	622	21	AAZ16219	Human colon cancer
556	24	0.5	624	21	AAZ16354	Human colon cancer
557	24	0.5	627	21	AAZ15999	Human colon cancer
558	24	0.5	628	21	AAZ69661	Human ERK activato
559	24	0.5	629	21	AAZ16300	Human colon cancer
560	24	0.5	630	20	AAZ36690	Mouse OHP106 prote
561	24	0.5	631	19	AAZ82801	ATG-1120 (allograf
562	24	0.5	631	21	AAZ16349	Human colon cancer
563	24	0.5	632	21	AAZ16494	Human colon cancer
564	24	0.5	632	21	AAZ80714	Human colon cancer
565	24	0.5	633	19	AAV40479	Human secreted pro
566	24	0.5	638	21	AAZ60768	Pig SENS ligand nu
567	24	0.5	639	21	AAZ95280	Corn cyclin-depend
568	24	0.5	639	21	AAZ98055	Human secreted pro
569	24	0.5	640	20	AAZ30333	DNA encoding a hum
570	24	0.5	643	20	AAZ84539	Human secreted pro
571	24	0.5	643	21	AAZ68086	Human secreted pro
572	24	0.5	644	21	AAZ57403	Hepatitis C virus
573	24	0.5	645	21	AAZ80022	Human colon cancer
574	24	0.5	649	21	AAZ58893	Human tumour suppr
575	24	0.5	656	20	AAZ41245	Human normal ovari
576	24	0.5	657	21	AAZ69500	Human secreted pro
577	24	0.5	657	21	AAZ16086	Human colon cancer
578	24	0.5	658	21	AAZ80732	Human colon cancer
579	24	0.5	660	19	AAZ27070	Double-stranded ad
580	24	0.5	664	21	AAZ26336	Human secreted pro
581	24	0.5	667	20	AAZ209052	Differentiation in
582	24	0.5	667	21	AAZ46304	Human spermatiat
583	24	0.5	668	20	AAZ06772	Marmoset sperm-spe
584	24	0.5	669	21	AAZ80698	Human colon cancer
585	24	0.5	685	20	AAZ27366	Human secreted pro
586	24	0.5	685	21	AAZ64992	Membrane-bound pro
587	24	0.5	685	22	AAZ44138	Human PRO820 (UNO5
588	24	0.5	686	20	AAZ61374	DNA encoding a hum
589	24	0.5	687	20	AAZ30410	DNA encoding a hum
590	24	0.5	688	17	AAZ16856	Integrin subunit b
591	24	0.5	690	21	AAZ79990	Human colon cancer
592	24	0.5	693	21	AAZ77669	Human PRO1286 cdNA
593	24	0.5	693	21	AAZ37065	Human PRO1286 (UNQ
594	24	0.5	693	22	AAZ54301	DNA encoding prote
595	24	0.5	693	22	AAZ91565	Human PRO1286 cdNA
596	24	0.5	698	21	AAZ52541	Human secreted pro
597	24	0.5	704	20	AAZ19425	Nicotiana tabacum
598	24	0.5	705	21	AAZ21389	Human low adenosin
599	24	0.5	705	21	AAZ35267	Human adenosine re
600	24	0.5	705	21	AAZ80590	Human colon cancer
601	24	0.5	706	20	AAZ87358	Human P2X3 puriner
602	24	0.5	706	21	AAZ64100	Human P2X-3 recept
603	24	0.5	708	20	AAZ61392	DNA encoding a hum
604	24	0.5	710	21	AAZ02121	Human colon cancer
605	24	0.5	712	19	AAZ59811	Human secreted pro
606	24	0.5	719	20	AAZ27374	Human secreted pro
607	24	0.5	722	21	AAZ44453	Arabidopsis thalia
608	24	0.5	727	21	AAZ97470	Human prostate can
609	24	0.5	731	16	AAZ099804	PR-1 like gene PR-
610	24	0.5	732	21	AAZ80613	Human colon cancer
611	24	0.5	732	22	AAZ44861	Human breast cancer
612	24	0.5	733	20	AAZ41378	Human normal uteru
613	24	0.5	735	19	AAZ29153	Mouse zins3 gene.
614	24	0.5	735	21	AAZ50153	Mouse insulin faml
615	24	0.5	738	20	AAZ30396	DNA encoding a hum
616	24	0.5	738	21	AAZ16171	Human prostate can
617	24	0.5	738	22	AAZ54313	DNA encoding prote
618	24	0.5	739	21	AAZ58615	Human PRO1283 prot
619	24	0.5	739	21	AAZ37071	Human PRO1283 (UNQ
620	24	0.5	741	21	AAZ97469	Human prostate can
621	24	0.5	745	17	AAZ47800	Oat-derived thioni
622	24	0.5	745	19	AAZ82783	Clone dn809_5 isol
623	24	0.5	745	21	AAZ87692	Human secreted pro
624	24	0.5	749	22	AAZ44900	Human breast cancer
625	24	0.5	750	21	AAZ59779	Human secreted pro
626	24	0.5	755	11	AAZ03970	Clone P25b. Mus m
627	24	0.5	756	21	AAZ77916	Human cancer assoc
628	24	0.5	767	16	AAZ85889	Wild-type bovine p
629	24	0.5	769	21	AAZ81222	Arabidopsis thalia
630	24	0.5	772	22	AAZ32766	Human secreted pro
631	24	0.5	774	17	AAZ10120	Vascular endothelli
632	24	0.5	774	18	AAZ85644	Antisense inhibito
633	24	0.5	774	18	AAZ79139	Human vascular end
634	24	0.5	774	19	AAZ15102	Human vascular end
635	24	0.5	774	19	AAZ95393	Human secreted pro
636	24	0.5	781	21	AAZ33349	Human prostate can
637	24	0.5	784	21	AAZ97286	Human secreted pro
638	24	0.5	785	20	AAZ24903	Human secreted pro
639	24	0.5	787	22	AAZ33240	Human secreted pro
640	24	0.5	788	11	AAZ001787	Partial sequence o
641	24	0.5	790	19	AAZ13117	Inserted DNA fragm
642	24	0.5	790	21	AAZ39903	DNA fragment from
643	24	0.5	797	18	AAZ02898	Human HMG1-C aberr
644	24	0.5	801	21	AAZ67695	Human ORFX ORF2550
645	24	0.5	805	20	AAZ52972	Human prostate tum
646	24	0.5	809	21	AAZ68125	Human secreted pro
647	24	0.5	813	21	AAZ87676	Human secreted pro
648	24	0.5	818	20	AAZ15208	Human gene express
649	24	0.5	821	21	AAZ02473	Human colon cancer
650	24	0.5	822	21	AAZ59202	Human secreted pro
651	24	0.5	823	21	AAZ23461	CDNA encoding huma
652	24	0.5	826	17	AAZ15229	Tumor necrosis fac
653	24	0.5	826	18	AAZ46332	TNF-RI-DD ligand p
654	24	0.5	827	21	AAZ78094	Human cancer assoc
655	24	0.5	830	18	AAZ62455	Lambda 5h-1 clone
656	24	0.5	831	20	AAZ89609	Human secreted pro
657	24	0.5	831	20	AAZ37428	Human secreted pro
658	24	0.5	832	20	AAZ30363	DNA encoding a hum
659	24	0.5	837	21	AAZ69637	Human secreted pro
660	24	0.5	839	22	AAZ86484	Maize Mac20 coding
661	24	0.5	841	21	AAZ52082	Pigment protein fr
662	24	0.5	843	20	AAZ06774	Marmoset sperm-spe
663	24	0.5	846	22	AAZ85089	Atherosclerosis-as
664	24	0.5	849	21	AAZ59394	Human secreted pro
665	24	0.5	855	21	AAZ21167	Human low adenosin
666	24	0.5	855	21	AAZ35045	Human adenosine re
667	24	0.5	859	20	AAZ97989	Human secreted pro
668	24	0.5	863	18	AAZ91302	Human H1075-1 secr

377	24	0.5	98	21	AAC21458	Human secreted pro	c 450	24	0.5	327	21	AAA31562	Plant microsatelli
378	24	0.5	98	21	AAC21773	Human secreted pro	451	24	0.5	327	21	AAA16539	Human colon cancer
379	24	0.5	100	18	AX83229	Breast cancer tumo	c 452	24	0.5	330	20	AAV83960	Bacterial artifici
380	24	0.5	100	19	AAV68827	DNA molecule encod	c 453	24	0.5	334	21	AAC19789	Human secreted pro
381	24	0.5	100	21	AAC80752	Human breast tumou	c 454	24	0.5	336	21	AAA16503	Human colon cancer
382	24	0.5	101	21	AAC18095	Human secreted pro	c 455	24	0.5	340	21	AAC98349	Human colon cancer
383	24	0.5	102	16	AAQ99624	Homology vector 52	456	24	0.5	342	21	AAA16337	Human colon cancer
384	24	0.5	105	21	AAQ60459	Humine factor V 5'	457	24	0.5	344	22	AAF67203	Novel human polynu
385	24	0.5	105	21	AAC29274	Human secreted pro	458	24	0.5	352	14	AAQ46676	Sequence Inserted
386	24	0.5	106	18	AAV00430	3' fragment of clo	c 459	24	0.5	352	21	AAF16488	Human prostate can
387	24	0.5	106	18	AAV46418	Ribosomal RNA DNA	460	24	0.5	352	21	AAA16524	Human colon cancer
388	24	0.5	108	16	AAQ99635	Homology vector 53	c 461	24	0.5	353	21	AAC77649	Human cancer assoc
389	24	0.5	110	20	AX35561	Secreted protein c	c 462	24	0.5	364	20	AAV87202	EST clone BN33. H
390	24	0.5	110	21	AAC13607	Human secreted pro	463	24	0.5	364	21	AAC93748	Cat flea hindgut a
391	24	0.5	121	21	AAC78441	Human cancer assoc	464	24	0.5	371	22	AAF68094	Human lung tumour
392	24	0.5	130	21	AAC16479	Human secreted pro	c 465	24	0.5	372	21	AAC99865	Human secreted pro
393	24	0.5	140	21	AAC29478	Human secreted pro	c 466	24	0.5	373	21	AAA40500	Human fetal brain
394	24	0.5	160	22	AAF24360	Retroviral recombi	467	24	0.5	373	22	AAF66909	Novel human polynu
395	24	0.5	164	21	AAF18008	Lung cancer associ	468	24	0.5	374	21	AAA16130	Human colon cancer
396	24	0.5	165	17	AAF34610	Probe for detectin	469	24	0.5	378	21	AAAD00146	TR12 related DNA-1
397	24	0.5	170	21	AAC13889	Human secreted pro	470	24	0.5	378	22	AAF66903	Novel human polynu
398	24	0.5	170	21	AAZ42381	Human 5' EST isola	c 471	24	0.5	386	21	AAC78443	Human cancer assoc
399	24	0.5	178	15	AAQ68946	Junction B of SfII	472	24	0.5	388	22	AAF65550	Novel human polynu
400	24	0.5	178	15	AAQ70553	Junction B of SfII	473	24	0.5	396	22	AAF94836	Human ovarian can
401	24	0.5	178	18	AAZ48503	Homology vector 50	474	24	0.5	396	22	AAF64568	Novel human polynu
402	24	0.5	178	20	AAZ81149	Sequence of juncti	475	24	0.5	403	21	AAA16024	Human colon cancer
403	24	0.5	178	21	AAC67864	FPV homology vecto	476	24	0.5	404	22	AAF66780	Novel human polynu
404	24	0.5	178	21	AAZ49297	FPV homology vecto	477	24	0.5	406	22	AAF54712	Nucleotide sequenc
405	24	0.5	178	22	AAF24361	Retroviral recombi	478	24	0.5	406	22	AAF66944	Novel human polynu
406	24	0.5	179	19	AAV11620	Homo sapiens adult	c 479	24	0.5	428	21	AAC55708	Human differential
407	24	0.5	182	15	AAQ70552	Junction A of SfII	c 480	24	0.5	433	18	AAV00428	3' fragment of clo
408	24	0.5	182	18	AAZ48502	Homology vector 50	c 481	24	0.5	433	21	AAC98176	Human colon cancer
409	24	0.5	182	20	AAZ81148	Sequence of juncti	c 482	24	0.5	433	22	AAC85068	Atherosclerosis-as
410	24	0.5	182	21	AAC67863	FPV homology vecto	c 483	24	0.5	437	18	AAZ72938	Phaffia glyceralde
411	24	0.5	182	21	AAC11449	Human secreted pro	484	24	0.5	438	14	AAQ46679	Sequence Inserted
412	24	0.5	182	21	AAZ49296	FPV homology vecto	c 485	24	0.5	441	20	AAZ09474	pl35-WT3 construct
413	24	0.5	183	14	AAQ60893	Human brain Expres	c 486	24	0.5	441	22	AAF75092	Human colon associ
414	24	0.5	190	18	AAZ73673	Competitor No. 1 u	487	24	0.5	450	21	AAA45071	Human secreted exp
415	24	0.5	190	18	AAZ73675	Competitor No. 3 u	c 488	24	0.5	460	20	AAV87401	EST clone CDI94
416	24	0.5	192	21	AAC13025	Human secreted pro	c 489	24	0.5	466	21	AAC94840	Cat flea hindgut a
417	24	0.5	196	20	AAZ21523	Progression suppre	490	24	0.5	470	19	AAV61312	cDNA sequence of p
418	24	0.5	198	22	AAF60573	Influenza neuramin	491	24	0.5	470	19	AAV58579	Prostate tumour sp
419	24	0.5	205	16	AAQ99629	Homology vector 53	492	24	0.5	470	21	AAA06342	Human immunogenic
420	24	0.5	207	14	AAQ93099	GFP - IE, part of	c 493	24	0.5	477	21	AAC93703	Human secreted pro
421	24	0.5	208	21	AAA42364	Human secreted exp	494	24	0.5	479	21	AAC80348	Human colon cancer
422	24	0.5	208	22	AAF24363	Retroviral recombi	c 495	24	0.5	485	21	AAC65037	Membrane-bound pro
423	24	0.5	217	21	AAC15434	Human secreted pro	c 496	24	0.5	485	22	AAF92079	Human PRO1003 cDNA
424	24	0.5	218	21	AAA50090	Arabidopsis herbic	c 497	24	0.5	485	22	AAF44183	Human secreted pro
425	24	0.5	220	18	AAZ73674	Competitor No. 2 u	c 498	24	0.5	493	21	AAC66424	Human secreted pro
426	24	0.5	224	21	AAA45317	Human secreted exp	499	24	0.5	498	21	AAC93800	Cat flea hindgut a
427	24	0.5	226	21	AAC14683	Human secreted pro	500	24	0.5	498	21	AAC94407	Cat flea hindgut a
428	24	0.5	230	21	AAC29579	Human secreted pro	c 501	24	0.5	506	19	AAV59746	Human secreted pro
429	24	0.5	236	21	AAC93680	Cat flea hindgut a	502	24	0.5	509	19	AAV61342	Extended cDNA seq
430	24	0.5	256	21	AAA16182	Human colon cancer	503	24	0.5	509	19	AAV58672	Prostate tumour sp
431	24	0.5	259	18	AAZ91303	Human H1075-1 secr	c 504	24	0.5	509	21	AAC59006	Human secreted pro
432	24	0.5	259	18	AAV00423	3' fragment of clo	505	24	0.5	509	21	AAA08435	Human immunogenic
433	24	0.5	263	18	AAZ88781	Malassezia fungus	c 506	24	0.5	517	22	AAAS4399	Plant defensin cod
434	24	0.5	263	20	AAV88738	EST clone HJ674	c 507	24	0.5	522	21	AAF12440	Aspergillus oryzae
435	24	0.5	277	21	AAA44945	Human secreted exp	508	24	0.5	523	14	AAQ46677	Sequence Inserted
436	24	0.5	285	19	AAV11623	Homo sapiens adult	509	24	0.5	526	16	AAQ86309	Wilson disease gen
437	24	0.5	288	20	AAV89510	EST clone COL151	c 510	24	0.5	534	21	AAA38006	UCK-1 nucleotide s
438	24	0.5	296	21	AAAL6014	Human colon cancer	511	24	0.5	537	21	AAZ57401	Hepatitis C virus
439	24	0.5	299	19	AAV69628	Human secreted pro	c 512	24	0.5	541	20	AAZ87415	Hepatocellular car
440	24	0.5	301	14	AAQ59562	Human brain Expres	c 513	24	0.5	543	21	AAC81743	Human secreted pro
441	24	0.5	304	20	AAZ21531	Progression elevat	c 514	24	0.5	546	22	AAF44946	Human breast cance
442	24	0.5	305	20	AAV86214	EST clone M338. H	c 515	24	0.5	549	19	AAV38759	cDNA nfSP17-549 en
443	24	0.5	307	22	AAF68866	Human lung tumour	516	24	0.5	549	19	AAV38760	Complementary stra
444	24	0.5	317	14	AAQ59920	Human brain Expres	c 517	24	0.5	550	21	AAZ52554	Human secreted pro
445	24	0.5	318	18	AAZ91312	Human J143-1 secre	518	24	0.5	553	19	AAV46447	Human chromosome 7
446	24	0.5	322	21	AAV00436	3' fragment of clo	c 519	24	0.5	556	19	AAV66764	Pathogen response
447	24	0.5	322	21	AAC98002	Human colon cancer	c 520	24	0.5	563	21	AAA50065	Prostate cancer as
448	24	0.5	322	21	AAAL6211	Human colon cancer	c 521	24	0.5	565	21	AAA04013	Potato infection i
449	24	0.5	322	21	AAA16357	Human colon cancer	c 522	24	0.5	565	21	AAA45987	Human metastatic m

c 231	25	0.5	2730	22	AAE29459	Human TANGO 332 cD	c 304	25	0.5	12827	19	AAV09036	Equine arthritis v
c 232	25	0.5	2738	19	AAV48220	APEG-1 5' vascular	c 305	25	0.5	13104	14	AAQ46852	Clone of recombin
c 233	25	0.5	2738	21	AA251046	2.7 kb fragment co	c 306	25	0.5	13206	11	AAQ05243	Rat acyl peptide h
c 234	25	0.5	2812	21	AAE16219	Human prostate can	c 307	25	0.5	15528	19	AAV09039	Equine arthritis v
c 235	25	0.5	2812	20	AAE27909	Canine B7-1 protei	c 308	25	0.5	16956	21	AAZ91923	Wild type (C57BL/6
c 236	25	0.5	2830	20	AAZ27910	Canine B7-1 gene c	c 309	25	0.5	20303	18	AAZ71699	Human deoxycytidyl
c 237	25	0.5	2840	21	AAZ78596	Human PRO4993 nucl	c 310	25	0.5	21742	21	AAZ20938	Human high affinit
c 238	25	0.5	2980	21	AAZ23444	CDNA encoding huma	c 311	25	0.5	21742	21	AAA34816	Human adenosine re
c 239	25	0.5	3046	19	AAZ29074	Rattus norvegicus	c 312	25	0.5	22540	21	AAC88442	Thiamine responsiv
c 240	25	0.5	3080	21	AAZ78415	Human secreted pro	c 313	25	0.5	24000	21	AAA88551	Human dual-specific
c 241	25	0.5	3114	18	AAV02800	Human RHAMM cDNA.	c 314	25	0.5	25871	21	AAA98888	Human genomic OCTN
c 242	25	0.5	3212	22	AAC84373	Nucleotide sequenc	c 315	25	0.5	26016	19	AAV57272	Human flavin-conta
c 243	25	0.5	3355	21	AAE16121	Human prostate can	c 316	25	0.5	26764	18	AAZ71696	Human deoxycytidyl
c 244	25	0.5	3413	21	AAZ51048	Mouse APEG-1 gene	c 317	25	0.5	33030	22	AAZ29337	Atopy related gene
c 245	25	0.5	3451	20	AAZ35701	CDNA encoding a pr	c 318	25	0.5	34488	22	AAZ97854	Human neuroblastom
c 246	25	0.5	3455	22	AAE32241	Mouse urinary tryp	c 319	25	0.5	35455	22	AAE54723	Nucleotide sequenc
c 247	25	0.5	3657	21	AAZ76218	Human ORFX ORF1773	c 320	25	0.5	39796	21	AAC61681	Nucleotide sequenc
c 248	25	0.5	3721	21	AAZ96343	CDNA encoding a no	c 321	25	0.5	45546	20	AAZ23520	Human kidney amino
c 249	25	0.5	3729	22	AAZ77688	Human wild-type Fc	c 322	25	0.5	49999	20	AAZ23891	Murine LOBO genom
c 250	25	0.5	3798	14	AAQ42424	ADA2 DNA. Synthet	c 323	25	0.5	49999	20	AAZ23896	Murine LOBO homolo
c 251	25	0.5	3839	22	AAZ32759	Human secreted pro	c 324	25	0.5	49999	20	AAZ23904	Human LOBO homolog
c 252	25	0.5	3842	20	AAZ23533	Tomato Xa21 clone	c 325	25	0.5	50000	21	AAA96364	Polymorphic repeat
c 253	25	0.5	3851	21	AAZ77558	Human ORFX ORF3113	c 326	25	0.5	72604	20	AAZ10752	Genomic sequence o
c 254	25	0.5	3901	21	AAZ33361	Human secreted pro	c 327	25	0.5	78925	21	AAC89888	Human FN gene. Ho
c 255	25	0.5	3938	20	AAZ05823	Rat pheromone rece	c 328	25	0.5	80595	20	AAZ83939	HC-contig derived
c 256	25	0.5	4066	19	AAZ59181	Human bak gene pro	c 329	25	0.5	81001	22	AAZ93035	Human apolipoprote
c 257	25	0.5	4131	22	AAZ77689	Human variant Fcep	c 330	25	0.5	81369	21	AAZ97997	Human T gene DNA.
c 258	25	0.5	4180	10	AAZ91773	Rat androgen recep	c 331	25	0.5	87350	18	AAZ83003	Human WRN genomic
c 259	25	0.5	4321	19	AAZ52853	Human eyal-c encod	c 332	25	0.5	106746	21	AAZ10225	Human PCTA-1 genom
c 260	25	0.5	4366	19	AAZ94063	Human growth hormo	c 333	25	0.5	117609	21	AAZ21435	Human receptor-rel
c 261	25	0.5	4436	9	AAZ81716	Human growth hormo	c 334	25	0.5	121162	21	AAC66548	Human kinasin-like
c 262	25	0.5	4450	19	AAZ15448	Arabidopsis PR-1 p	c 335	25	0.5	138169	21	AAA34791	Human adenosine re
c 263	25	0.5	4698	13	AAQ32370	PLA genomic DNA.	c 336	25	0.5	141589	21	AAZ20913	Human ELAM-1 polyn
c 264	25	0.5	4698	15	AAQ72475	Entire PLA genomic	c 337	25	0.5	141589	21	AAZ21127	Human low adenosin
c 265	25	0.5	4698	16	AAZ01176	P815A antigen prec	c 338	25	0.5	141589	21	AAZ21152	Human low adenosin
c 266	25	0.5	4698	17	AAZ14692	MAGE genomic DNA.	c 339	25	0.5	141589	21	AAA35005	Human adenosine re
c 267	25	0.5	4698	20	AAZ84111	PLA genomic sequen	c 340	25	0.5	141589	21	AAA35030	Human adenosine re
c 268	25	0.5	4749	18	AAZ59230	FSPI promoter link	c 341	25	0.5	146981	21	AAZ21442	Human ELAM-1 polyn
c 269	25	0.5	5235	21	AAZ76134	Human OREX ORF1689	c 342	25	0.5	151826	21	AAZ22291	BAC containing rep
c 270	25	0.5	5261	16	AAQ99602	Human Ah receptor	c 343	25	0.5	160552	22	AAZ02697	Human glycosyl sul
c 271	25	0.5	5261	18	AAZ85436	Human Ah-receptor	c 344	25	0.5	162450	21	AAZ86967	Human glycosyl sul
c 272	25	0.5	5301	18	AAZ93627	Human Wnt-1 gene	c 345	25	0.5	162450	21	AAZ86967	Human glycosyl sul
c 273	25	0.5	5305	20	AAZ28289	Human metastasis-a	c 346	25	0.5	209273	21	AAZ21437	Retinoblastoma bin
c 274	25	0.5	5555	15	AAQ64898	Rat neuronal immed	c 347	25	0.5	240825	22	AAZ24497	Human factor-3 gene
c 275	25	0.5	5555	17	AAZ18533	HER4 with alternat	c 348	24	0.5	38	21	AAZ43893	M. tuberculosis rp
c 276	25	0.5	5607	21	AAZ88884	Receptor tyrosine	c 349	24	0.5	40	13	AAZ25027	Anti-sense oligonu
c 277	25	0.5	5607	21	AAZ47788	Human Wnt-1 gene	c 350	24	0.5	40	13	AAZ25027	Sequence of primer
c 278	25	0.5	5836	16	AAQ96296	Human prostrate-sp	c 351	24	0.5	47	20	AAZ01041	Probe for human PG
c 279	25	0.5	5836	18	AAZ84444	Human prostrate-sp	c 352	24	0.5	48	16	AAZ01041	Trypsin inhibitory
c 280	25	0.5	5836	21	AAZ60485	5' flanking region	c 353	24	0.5	48	16	AAZ01041	Primer SINKball700
c 281	25	0.5	5836	21	AAZ60485	Nucleotide sequenc	c 354	24	0.5	48	17	AAZ35073	Sindbis PCR primer
c 282	25	0.5	5836	21	AAZ39130	Human prostrate spe	c 355	24	0.5	48	17	AAZ35073	Sindbis PCR primer
c 283	25	0.5	5836	21	AAZ99932	DNA sequence of co	c 356	24	0.5	48	19	AAZ42384	Reverse PCR primer
c 284	25	0.5	5888	20	AAZ24013	Human GDNF fragmen	c 357	24	0.5	48	20	AAZ70704	Reverse PCR primer
c 285	25	0.5	6083	21	AAA51398	Chromosome 16q tum	c 358	24	0.5	48	21	AAZ92785	Sindbis basic vect
c 286	25	0.5	6792	21	AAZ94062	Aryl hydrocarbon r	c 359	24	0.5	48	21	AAZ92912	Sindbis basic vect
c 287	25	0.5	7063	20	AAZ81916	CDNA encoding huma	c 360	24	0.5	53	21	AAZ18812	Human secreted pro
c 288	25	0.5	7063	20	AAZ81788	CDNA encoding huma	c 361	24	0.5	54	21	AAZ18812	Human secreted pro
c 289	25	0.5	7397	19	AAZ60578	Human tumour suppr	c 362	24	0.5	56	17	AAZ17032	Human mitochondria
c 290	25	0.5	7452	16	AAQ85594	Human glycoprotein	c 363	24	0.5	60	18	AAZ69401	PCR primer AL1. S
c 291	25	0.5	7452	21	AAZ61217	DNA encoding huma	c 364	24	0.5	60	21	AAZ14231	Human secreted pro
c 292	25	0.5	7452	22	AAZ90646	Human secreted pro	c 365	24	0.5	60	22	AAZ62064	Human secreted pro
c 293	25	0.5	8365	20	AAZ20056	Plasmodium falcipa	c 366	24	0.5	67	21	AAZ14340	Human secreted pro
c 294	25	0.5	10120	20	AAZ23683	Human DKC1 DNA fra	c 367	24	0.5	69	21	AAZ14340	Human secreted pro
c 295	25	0.5	10120	20	AAZ23683	Human DKC1 DNA fra	c 368	24	0.5	72	21	AAZ27585	Human secreted pro
c 296	25	0.5	11298	18	AAZ86756	Human high affinit	c 369	24	0.5	73	21	AAZ12555	Human secreted pro
c 297	25	0.5	11298	19	AAZ54661	Human beta subunit	c 370	24	0.5	77	21	AAZ14910	Human secreted pro
c 298	25	0.5	11298	21	AAZ20937	Human high affinit	c 371	24	0.5	83	21	AAZ13898	Human secreted pro
c 299	25	0.5	11298	22	AAZ34815	Human adenosine re	c 372	24	0.5	85	19	AAZ32418	Homo sapiens clone
c 300	25	0.5	11298	22	AAZ92144	Human IGERB gene S	c 373	24	0.5	87	21	AAZ23541	Human secreted pro
c 301	25	0.5	11357	14	AAQ51024	Human FCER1 beta c	c 374	24	0.5	90	21	AAZ22118	Human secreted pro
c 302	25	0.5	11967	22	AAZ97863	Human neuroblastom	c 375	24	0.5	92	18	AAZ88439	Padlock probe alph
c 303	25	0.5	12117	21	AAA96368	Polymorphic repeat	c 376	24	0.5	98	20	AAZ00174	Porcine reproducti

c 85	25	0.5	294	20	AAV82725	Partial sequence o	c 158	25	0.5	1149	21	AAA58565	DNA encoding human
c 86	25	0.5	294	21	AAA89542	Xyloglucan endoglu	c 159	25	0.5	1173	21	AAA63941	Nucleotide sequenc
c 87	25	0.5	294	21	AAA90868	Xyloglucan endoglu	c 160	25	0.5	1192	21	AAA51406	Chromosome 16q tum
c 88	25	0.5	294	22	AAF77876	Partial XGU endogl	c 161	25	0.5	1197	17	AAAT32429	Wasp venom BrhTX-1
c 89	25	0.5	294	22	AAF79737	Aspergillus aculea	c 162	25	0.5	1197	19	AAV17146	Insecticidal toxin
c 90	25	0.5	294	22	AAF81169	Aspergillus aculea	c 163	25	0.5	1201	17	AAAT32430	Wasp venom BrhTX-1
c 91	25	0.5	294	22	AAF75790	Partial XGU endogl	c 164	25	0.5	1210	21	AAZ50064	Impatiens balsamia
c 92	25	0.5	294	22	AAF77567	A aculeatus xylogl	c 165	25	0.5	1240	20	AAZ41296	Human normal ovary
c 93	25	0.5	294	22	AAF59749	3' portion of cDNA	c 166	25	0.5	1259	19	AAV52846	Human eyal gene co
c 94	25	0.5	327	18	AAAT88091	Human secreted pro	c 167	25	0.5	1284	21	AAV59950	Human secreted pro
c 95	25	0.5	327	18	AAV02166	Human secreted pro	c 168	25	0.5	1285	20	AAV84620	Human TNF11 coding
c 96	25	0.5	387	22	AAV65820	Novel human polynu	c 169	25	0.5	1285	21	AAV84620	Human TNF11 coding
c 97	25	0.5	390	21	AAA16227	Human colon cancer	c 170	25	0.5	1302	12	AAQ11112	MHC class II anti
c 98	25	0.5	416	21	AAV99917	Human secreted pro	c 171	25	0.5	1321	21	AAV93361	Arabidopsis thalia
c 99	25	0.5	419	21	AAV74314	Human secreted pro	c 172	25	0.5	1322	20	AAV04402	Human secreted pro
c 100	25	0.5	434	21	AAV74424	Human secreted pro	c 173	25	0.5	1324	20	AAV04402	Arabidopsis thalia
c 101	25	0.5	443	21	AAV75749	cDNA encoding a de	c 174	25	0.5	1337	20	AAV04345	Human secreted pro
c 102	25	0.5	484	22	AAV28388	Anopheles albimanu	c 175	25	0.5	1348	19	AAV34289	Human secreted pro
c 103	25	0.5	484	22	AAV28388	Barley Clone F. H	c 176	25	0.5	1353	21	AAV18255	Lung cancer associ
c 104	25	0.5	499	21	AAZ97350	Human prostate can	c 177	25	0.5	1359	21	AAV98822	Human secreted pro
c 105	25	0.5	500	21	AAV93907	Cat flea hindgut a	c 178	25	0.5	1396	16	AAQ98498	Human 50 kDa dystr
c 106	25	0.5	528	18	AAV02891	Human HMGI-C aberr	c 179	25	0.5	1396	18	AAV97324	Human (50 kDa) dys
c 107	25	0.5	528	18	AAV02891	Human secreted pro	c 180	25	0.5	1442	21	AAV80566	Human secreted pro
c 108	25	0.5	551	21	AAV21818	Human breast and o	c 181	25	0.5	1444	21	AAV77169	Human ORFX ORF2724
c 109	25	0.5	553	19	AAV34276	Human secreted pro	c 182	25	0.5	1447	13	AAQ30002	HCNP precursor gen
c 110	25	0.5	554	22	AAV33138	Human secreted pro	c 183	25	0.5	1447	15	AAV58686	Human hippocampal
c 111	25	0.5	557	19	AAV34223	Human secreted pro	c 184	25	0.5	1467	21	AAZ52508	Human secreted pro
c 112	25	0.5	569	20	AAV27389	Human secreted pro	c 185	25	0.5	1487	20	AAV00715	Human secreted pro
c 113	25	0.5	571	20	AAV06228	Human secreted pro	c 186	25	0.5	1487	21	AAV297117	Human secreted pro
c 114	25	0.5	593	21	AAV26329	Human secreted pro	c 187	25	0.5	1492	21	AAV51403	Chromosome 16q tum
c 115	25	0.5	604	21	AAV59787	Human secreted pro	c 188	25	0.5	1494	11	AAQ03369	Sequence encoding
c 116	25	0.5	611	21	AAV16388	Human colon cancer	c 189	25	0.5	1496	10	AAV92386	DNA coding for the
c 117	25	0.5	615	21	AAV16139	Human colon cancer	c 190	25	0.5	1512	18	AAV72172	Alzheimer's diseas
c 118	25	0.5	615	18	AAV02881	Human HMGI-C aberr	c 191	25	0.5	1534	22	AAV23910	Human secreted pro
c 119	25	0.5	617	18	AAV02870	Human HMGI-C/PSAFP	c 192	25	0.5	1558	20	AAV84453	Human secreted pro
c 120	25	0.5	618	21	AAV03065	Arabidopsis thalia	c 193	25	0.5	1563	20	AAV53007	Human prostate tum
c 121	25	0.5	625	21	AAV32915	Arabidopsis thalia	c 194	25	0.5	1570	21	AAV37103	Human secreted pro
c 122	25	0.5	636	21	AAV59368	Human secreted pro	c 195	25	0.5	1570	22	AAV54409	Human secreted pro
c 123	25	0.5	653	19	AAV52938	Pig F7CD antigen e	c 196	25	0.5	1622	21	AAV59771	Human secreted pro
c 124	25	0.5	658	21	AAV16322	Human colon cancer	c 197	25	0.5	1645	21	AAV21831	Human breast and o
c 125	25	0.5	689	18	AAV02893	Human HMGI-C aberr	c 198	25	0.5	1660	19	AAV48147	Nicotianamine amin
c 126	25	0.5	707	22	AAV27958	Human dual specifi	c 199	25	0.5	1671	22	AAV33038	Human secreted pro
c 127	25	0.5	715	20	AAV37514	Human secreted pro	c 200	25	0.5	1682	21	AAV18279	Lung cancer associ
c 128	25	0.5	725	21	AAV77602	Human ORFX ORF3157	c 201	25	0.5	1693	22	AAV63803	Human secreted pro
c 129	25	0.5	726	20	AAV03016	Human IL-1ra BAC c	c 202	25	0.5	1831	21	AAV79902	Human secreted pro
c 130	25	0.5	731	13	AAQ31693	RsaI restriction f	c 203	25	0.5	1839	21	AAV78067	Human cancer assoc
c 131	25	0.5	732	21	AAV52519	Human secreted pro	c 204	25	0.5	1894	20	AAV24072	Human hRNP G DNA
c 132	25	0.5	740	20	AAV52980	Human prostate tum	c 205	25	0.5	1927	21	AAV79919	Human secreted pro
c 133	25	0.5	757	21	AAV79355	Plant viral move	c 206	25	0.5	1955	21	AAV93102	Human secreted pro
c 134	25	0.5	815	20	AAV52884	Human prostate tum	c 207	25	0.5	1986	21	AAV93102	Human secreted pro
c 135	25	0.5	825	18	AAV86069	Partial cDNA clone	c 208	25	0.5	2020	22	AAV02199	Chondrus crispus s
c 136	25	0.5	825	18	AAV02157	Human secreted pro	c 209	25	0.5	2058	14	AAQ41227	Clone GP46 encodin
c 137	25	0.5	825	21	AAV040502	Human adult blood	c 210	25	0.5	2085	21	AAV93408	Human secreted pro
c 138	25	0.5	860	21	AAV79359	Plant viral move	c 211	25	0.5	2094	20	AAV19489	Human secreted pro
c 139	25	0.5	870	21	AAV99870	Human secreted pro	c 212	25	0.5	2094	21	AAV99870	Human secreted pro
c 140	25	0.5	876	21	AAV63439	Human secreted pro	c 213	25	0.5	2114	20	AAV19958	Rat pango-76 encod
c 141	25	0.5	880	21	AAV21883	Human breast and o	c 214	25	0.5	2139	21	AAV79860	Human secreted pro
c 142	25	0.5	890	21	AAV66182	Human trypsin hl g	c 215	25	0.5	2201	21	AAV43777	Human fetal brain
c 143	25	0.5	890	21	AAV292417	cDNA encoding huma	c 216	25	0.5	2202	21	AAV52286	Maize replicon
c 144	25	0.5	903	21	AAV64431	DNA encoding a hum	c 217	25	0.5	2226	15	AAV72270	Rat MR22 serotonin
c 145	25	0.5	911	21	AAV59124	Human secreted pro	c 218	25	0.5	2238	21	AAV59408	Human secreted pro
c 146	25	0.5	971	21	AAV58883	Human tumour suppr	c 219	25	0.5	2270	20	AAV24865	Human secreted pro
c 147	25	0.5	979	19	AAV69626	Human secreted pro	c 220	25	0.5	2328	18	AAV95227	Human islet cell a
c 148	25	0.5	989	20	AAV32618	Human breast tumou	c 221	25	0.5	2387	21	AAV23374	Human secreted pro
c 149	25	0.5	1013	21	AAV25487	Canine Flt-3 ligand	c 222	25	0.5	2394	22	AAV585095	Atherosclerosis-as
c 150	25	0.5	1013	21	AAV25488	Canine Flt-3 ligand	c 223	25	0.5	2402	19	AAV07202	Mouse epandymin-11
c 151	25	0.5	1024	21	AAV297148	Human prostate can	c 224	25	0.5	2432	20	AAV00650	Human secreted pro
c 152	25	0.5	1027	22	AAV89255	Mouse canello 1 (M	c 225	25	0.5	2435	20	AAV00707	Human secreted pro
c 153	25	0.5	1040	18	AAV02818	Human RHAMM genomi	c 226	25	0.5	2438	22	AAV92529	Rat T2R02 nucleoti
c 154	25	0.5	1054	21	AAV74245	Human secreted pro	c 227	25	0.5	2574	21	AAV27982	Human homology to
c 155	25	0.5	1076	21	AAV77715	Human cancer assoc	c 228	25	0.5	2599	21	AAV76501	Human ORFX ORF2056
c 156	25	0.5	1130	21	AAV39064	Human secreted pro	c 229	25	0.5	2625	21	AAV16658	Human secreted pro
c 157	25	0.5	1138	21	AAV58267	Rice cellulose syn	c 230	25	0.5	2628	18	AAV01060	Human PKR gene. H

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 16:19:42 ; Search time 299.68 Seconds
(without alignments)
11044.012 Million cell updates/sec

Title: US-09-512-581-1
Perfect score: 5271
Sequence: 1 ccggagagcccgagtgag.....gtataaaaaaaaaaaaaa 5271

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :	N_Geneseq_0601:*
1:	/SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	5271	100.0	5271	Human androgen shu
2	5271	100.0	5355	Human androgen shu
3	407	7.7	439	EST clone DY17. H
4	348	6.6	738	Human gene express
5	292	5.5	295	Human secreted pro
6	268	5.1	351	Human 5' EST sola
7	219	4.2	284	Human gene express
8	131	2.5	405	Human gene express
9	120	2.3	161	Human gene signatu
10	99	1.9	150	Human secreted pro
11	99	1.9	150	Human biallelic po

Human colon cancer	530	21	AAZ80598
Polynucleotide seq	421	20	AAZ21198
EST clone BP810	489	20	AAV87322
Nucleotide sequenc	14784	21	AAA64141
Sequence of a COSM	45265	21	AAZ46508
Human secreted pro	60	21	AAC23080
Human TAC1-ligand	1057	22	AAZ02007
Human secreted pro	680	21	AAC74320
Gastric cancer ass	704	20	AAZ40124
Human secreted pro	845	21	AAZ59318
Human secreted pro	1352	20	AAZ98002
Human secreted pro	1358	21	AAC59286
Human secreted pro	1419	21	AAC59246
Human breast tumou	1776	20	AAZ33581
Human PRO1082 nucl	2477	20	AAZ34387
Human PRO1082 (UMG	2477	21	AAZ78579
Human IL-1ra BAC c	3076	20	AAZ02987
Glucanase II gene	3447	19	AAV16444
Human IL4Ralpha qe	28690	22	AAZ57718
Genomic sequence o	43069	21	AAZ36335
Human Hscd66 intro	128	21	AAZ00579
Junction A of Sf11	182	15	AAQ68945
EST clone CO209	293	20	AAV89475
3' portion of cDNA	394	18	AAT88088
Human secreted pro	394	18	AAV02161
Human secreted pro	451	21	AAC93441
Human HMG1-C aberr	495	18	AAV02892
Human colon cancer	535	21	AAA16380
Murine glycosylati	635	16	AAQ75445
Murine glycosylati	723	18	AAT42755
Great tit CHD-W qe	794	21	AAC79751
Human secreted pro	807	17	AAT27152
Human Machado-Jose	1067	21	AAZ52545
Lung cancer associ	1216	21	AAZ18145
Human secreted pro	1346	21	AAC79040
Cucumber mosaic v1	1423	10	AAZ90249
Cucumber mosaic v1	1426	18	AAT72272
Human P2 DNA. Hom	1459	21	AAZ27155
Human secreted pro	1504	21	AAC93333
Human prostate can	1573	22	AAZ72752
Exemplary human P2	1657	21	AAZ27156
Human Machado-Jose	1776	20	AAZ23425
Human secreted pro	2118	21	AAC59907
Human adenosine A3	2470	21	AAZ20859
Human adenosine A3	2470	21	AAZ20870
Human adenosine re	2470	21	AAZ34737
Human adenosine re	2588	21	AAZ34748
Human ORFX ORF2311	2923	20	AAZ04326
Human secreted pro	3350	17	AAZ08553
Oncogene R-ras mut	3744	21	AAZ27157
Human P2 genomic D	5318	22	AAZ5961
Human eukaryotic 1	11395	21	AAZ21440
Human adenosine A3	9999	20	AAZ23903
Human LOBO homolog	72604	20	AAZ10752
Genomic sequence o	90050	21	AAZ91925
Wild type (C57BL/6	117609	21	AAZ21435
Human receptor-rel	121162	21	AAZ66548
Human kinesin-like	48	20	AAZ57022
WO9923258 oligonuc	54	15	AAQ63929
Synthetic Cholera	55	18	AAT72962
Omega probe P302 u	75	17	AAT17036
Human mitochondria	121	19	AAV54078
Nucleotide sequenc	164	21	AAC16464
Human secreted pro	186	21	AAC28130
Human secreted pro	226	17	AAT43285
Promoter for infec	232	21	AAC20507
Human secreted pro	254	21	AAC11615
Human secreted exp	265	21	AAA42328
EST clone CW762	283	20	AAV89992
EST clone DM360	284	20	AAV90426
Endoglucanase type	294	15	AAQ67025

AUTHORS

Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
Y., Gu,Y., Chen,Z., and Han,Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source

1..608
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKATF08"
/clone_lib="GK"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT

181 a 113 c 125 g 189 t

ORIGIN

Query Match

Best Local Similarity 2.9%; Score 151; DB 32; Length 608;

Matches 351; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 528 accagctatcacagaaccttatttttcagttataaacaatggccacatcagaagtcacat 587

Db 80 ACCAGCTATACAGAACCCTTATTTTCAGTTATAAACAATGGCCACATCAAGAGTCCAT 139

QY 588 atgcacatggtagaccttactagctctattttgtgaagtgatcacagtgctccaggag 647

Db 140 ATGCACATGTTAGACCTTATGAGCTCTATTATTGTGAAGTGATACAGTGTCTCAGGAG 199

QY 648 ctttggatacggtttttagtaaatctggtacctgctcataagaatttaacaagcaagca 707

Db 200 CTTTGGATACGGATTAGTAATCTGCTACCTGCTCATAGAAGTATTAAGCAAGCAAGCA 259

QY 708 tatgaattggcaaggcttactgaagagagacagctcaagctattgagccatatatacc 767

Db 260 TATGATTTGGCAAGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACC 319

QY 768 acttttttaacaggtctctgatgtcttgggaaacatctatcagcgattgtcagagcat 827

Db 320 AATTTTTTATCAGGATCTGATGCTTTGGGAAAACATCTATCAGCGATTGTCTCAGGAT 379

QY 828 gtctttgacttaatttggagctctacaatatattgatagtcatttgcgtgctctcg 882

Db 380 GTCTTTGACTTAATTTGGAGCTCTACAATATTGATAGTCATTGTGCTGCTCTCTG 434

RESULT 50

AL121425

LOCUS

DEFINITION DKFZP762H077.r1 762 bp mRNA EST 25-FEB-2000

DKFZP762H077.5, mRNA sequence.

ACCESSION AL121425

VERSION AL121425.1

KEYWORDS GI:5927426

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 167)

Ottewaeldeir,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and

Wiemann,S.

TITLE

JOURNAL

COMMENT

Contact: Ottewaeldeir B

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by Medigenomix (Martinsried/Germany) within the cDNA

sequencing consortium of the German Genome Project. No sl sequence

available.

This clone (DKFZP762H077) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source

1..167
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP762H077"
/clone_lib="762 (synonym: hmel2)"
/tissue_type="melanoma (MeWo cell line)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT

72 a 14 c 19 g 62 t

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 147; DB 105; Length 167;

Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5102 tggcaatgaaaattttaagaagaagatttaaaagtattttaattttaagaagtgtgttat 5161

Db 1 TGGCAATGAAATTTTAAAGAAGAAGATTTAAAGTATTTTAAAGAGTGTGTAT 60

QY 5162 aaaataatgtactgaattctttatccccattttatccatcttccagtttttatttaattac 5221

Db 61 AAATAATGTAATCTTTTATCCCATTTTATCATCTTTCAGTTTTTATTATTAATCTAC 120

QY 5222 tgatcaataaaattctgttaattgaa 5248

Db 121 TGTATCAATAAAATCTGTAATTGAA 147

Search completed: September 25, 2001, 20:54:46

Job time: 22799 sec


```

Db      6 TGA 4

RESULT 45
AA687846
LOCUS   AA687846
DEFINITION nv09a11.s1 NCI_CGAP_Pr-22 Homo sapiens cDNA clone IMAGE:1219676 3',
mRNA sequence.
ACCESSION AA687846
VERSION   AA687846
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapb-r@mail.nih.gov
          Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
          cDNA Library Preparation: M. Bento Soares, Ph.D.
          DNA Sequencing by: Greg Lennon, Ph.D.
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES             Location/Qualifiers
     source           1..421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1219676"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT73 vector. Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   128 a   71 c   69 g   153 t
ORIGIN

Query Match      3.4%; Score 181; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.3e-68;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1805 gcgtgaataactaagaagttgggcaaccaccccaacagccctacaaatcttctcggaaat 1864
Db 241 GCGTGAATAACTAAGAAGTTGGGCAACCCCAACACGCCCTACAAATCCTTCTCTGGAAAT 300
QY 1865 gatacaattctcttgagaggatgacacctgtgcacatagataccgaatctatcagtgc 1924
Db 301 GATCAAGTTTCTTGGAGGAGGATAGGACCTGTGCACATAGATACCGAATCTATCAGTGC 360
QY 1925 tcttattaacaagtgaacaaatcaatagatggcaacgacgagatgatgaagatgagggtgt 1984
Db 361 TCTTATTAAACAAGTGAACAAATCAATAGATGGGAACAGCAGCATGATGAAGATGAGGGTGT 420
QY 1985 t 1985
Db 421 t 421

```

```

RESULT 46
AI140791/C
LOCUS   AI140791
DEFINITION AI140791 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:1566224 3', mRNA sequence.
ACCESSION AI140791
VERSION   AI140791
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 306)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapb-r@mail.nih.gov
          This clone is available royalty-free through LLNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert length: 1585 Std Error: 0.00
          Seq primer: -40ml3 fwd. ET from Amersham
          High quality sequence stop: 230.

FEATURES             Location/Qualifiers
     source           1..306
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1566224"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and clones: Soares NBHF pool 1:
                     309384-310919, 323208-325895 Soares NB2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares NB2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NBHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NBHOT
                     pool 1: 723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."

BASE COUNT   110 a   43 c   39 g   114 t
ORIGIN

Query Match      3.1%; Score 164; DB 16; Length 306;
Best Local Similarity 99.5%; Pred. No. 6.7e-61;
Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5003 tgataggaggaacaccatgcaaatgtgaaatagctcgaatctcttgattactttaca 5062
Db 272 TGATAGGAGGACACCATGCAAAATGTGAAATAGTCTCGAAGTCTCTGGATTACTTTACA 213
QY 5063 cctcagattgattgtcccaagaattttctgcctttcatgccaatgaaatttaagaa 5122
Db 212 CCTCAGTATTGATTGTGCCAGAAATTTTGGCCCTTTCATGCAATGAAATTTAAGAA 153
QY 5123 gaaagattaaagcatttttaatttaagagagtgtgtgtataaaaaataagtactgaattctt 5182
Db 152 GAAAGATTAAAGTATTATTTTAAAGAGTGTGTATAAAATAATGTACTGAATTTCTT 93
QY 5183 tatccattttatcatctcttcagtttttattat 5217
Db 92 TATCCCATTTTATCATCTCTTCAGTTTTTTTATTAAT 58

```



```

RESULT 43
AA747568      249 bp      mRNA      EST      22-JAN-1998
LOCUS      nx92c01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269696 3',
DEFINITION      mRNA sequence.
ACCESSION      AA747568
VERSION      AA747568.1 GI:2787526
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 249)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbrrp/image/image.html
Insert Length: 532 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 190.
Location/Qualifiers
1. .249
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1269696"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dt) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTCTTTT-3',
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      61 a      33 c      46 g      109 t
ORIGIN

Query Match      3.8%; Score 198; DB 11; Length 249;
Best Local Similarity 99.6%; Pred. No. 8.2e-76;
Matches 248; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4666 ttttttaagtgttcgatttcgaagtcgttatagctttttatctgcggctttaact 4725
Db 1 TTTTAAAGTTTCTGATTTCTGAAGTGCCTGTATAGCTTTTATCTCGGCTTTAAACT 60

QY 4726 gacagtaccgcagctgtttatggatctattgttgaaagaatttggatagatct 4785
Db 61 GACAGTACCCGACTGTTATTCGATCTATTGATTGAAAGAAATTTGTAGGATAGATCT 120

QY 4786 taagcagtaattctgctgattgttattgtattttctgcaatttttactgtgaaaaaaa 4845
Db 121 TAAGCAGTAATCTGTCAGTGTGTTGTTATTTGTTATCTCTGCAATTTTCTGCAAAAAA 180

QY 4846 ttgttttcaacaattgggtgcatttttcttgatgctcactatttggagagtaaatgg 4905

```

```

Db 181 TTTGTTTCAACAATTTGGTGTCTATTTTCTGATGTCACATTTTGTGGAGAGTAAATGG 240
QY 4906 tctcttccc 4914
Db 241 TCTCTTCCC 249

RESULT 44
AI572349/c 186 bp      mRNA      EST      13-APR-1999
LOCUS      te40c12.x1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:2089174
DEFINITION      3', mRNA sequence.
ACCESSION      AI572349
VERSION      AI572349.1 GI:4535723
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 706 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 184.
Location/Qualifiers
1. .186
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2089174"
/clone_lib="Soares_NhMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      78 a      23 c      21 g      64 t
ORIGIN

Query Match      3.5%; Score 183; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 3.1e-69;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5069 tattgattgtcccgagaattttctgccccttcattggccaatgaaattttaagaagaaga 5128
Db 186 TATTGATTTGTCCCGAATTTTCTGGCCTTTTCATGCAATGAAATTTTAAAGAAGAAGA 127

QY 5129 tttaaagtatttaatttaaagagtggttatataaaatgaatcttctatccc 5188
Db 126 TTTAAGATATTTAATTTTAAGAGTGTGTATATAAATATGATCTGATCTTTATCCC 67

QY 5189 atttatcatcctttcagttttttatttaactctactgtatcaataaaattctgtaattgaa 5248
Db 66 ATTTTATCATCCTTCTCAGTTTTATTAACTACTGTATCAATAAATAATCTGTAAATTGAA 7

QY 5249 tga 5251

```



```

FEATURES
source
Location/Qualifiers
1. .446
/organism="Homo sapiens"
/db_xref="GDB:3848693"
/db_xref="taxon:9606"
/clone="IMAGE:220700"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: pT73D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 118 a 69 c 97 g 153 t 9 others
ORIGIN

Query Match 4.4%; Score 231; DB 158; Length 446;
Best Local Similarity 99.6%; Pred. No. 2.7e-90;
Matches 281; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4491 cacaaatggagctgtgaagagtgagcagttggaccttacttggtagcccccatacatt 4550
Db 1 CACAAAATGGAGCTGCTGAAGAGTGGACAGTTGGACCTTACTTTGGTGCACCCATACATT 60

QY 4551 tgtgtcacatgctttagccatacacatggttaacattgactatgagcttctgtgaagtg 4610
Db 61 TGTGTGCATGCTTTAGCCATACACATGTTAATGACTATGAGTCTTGTGAAGTG 120

QY 4611 taatgtcgatggtctatgtagacataaagaagaactgttaatatctttttttttt 4670
Db 121 TAATGTGCGATGGCTATGTAGACATAAAGAAGAACTGTGAATATATCTTTTCTTTT 180

QY 4671 ttaatgttctgtattctgaagtctgtatagcttttatcttgcggtttaaacctgacag 4730
Db 181 TTAATGTTCTGTATTCNGAAGTGTGTATAGCTTTTATCTGCGGCTTTAAACTGACAG 240

QY 4731 taccgcagctgttattggatctattgattgaaagaatttg 4772
Db 241 TACCGACTGTTTATGGATCTATTGATTGAAAGAAATTG 282

RESULT 38
BF222521 464 bp mRNA EST 09-NOV-2000
LOCUS BF222521
DEFINITION 7p53q08.x1 NCI-CCAP-Pr28 Homo sapiens cDNA clone IMAGE:3649719 3'
Similar to TR:Q9Y451 Q9Y451 ANDROGEN-INDUCED PROSTATE PROLIFERATIVE
SHUTOFF ASSOCIATED PROTEIN. ;, mRNA sequence.
ACCESSION BF222521
VERSION BF222521.1 GI:11129698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS Adams,M.D., Kerkavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gockayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

```

```

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
info@image.llnl.gov
High quality sequence stop: 412.
FEATURES
source
Location/Qualifiers
1. .464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3649719"
/clone_lib="NCI-CCAP-Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CCAP-Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 142 a 82 c 78 g 162 t
ORIGIN

Query Match 4.2%; Score 222; DB 146; Length 464;
Best Local Similarity 100.0%; Pred. No. 2.4e-86;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1805 gcgtgaataactaagaagttggcaccacacagcctacaaatccttctcctggaaat 1864
Db 239 GCGTGAATAACTAAGAAGTTGGCAACCCCAACAGCCTACAATCCTTCTTGGAAAT 298

QY 1865 gatcaagttctcttgagagagatagcacctgtgcacatagatcaccaaatctcagtcg 1924
Db 299 GATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGC 358

QY 1925 tcttataaacaagtgaacaaatcaatagatggaacagcagatgatgaagatgaggtgt 1984
Db 359 TCTTATTAAACAAGTGAACAAATCAATAGATGGACAGCAGATGATGAAGATGAGGGTGT 418

QY 1985 tccaactgatcaagccatcagagcaggtcttgaactgcctaa 2026
Db 419 TCCAACATGATCAAGCCATCAGAGCAGGTCTTTGAACCTGCTTAA 460

RESULT 39
AA333118 271 bp mRNA EST 21-APR-1997
LOCUS AA333118
DEFINITION EST37155 Embryo, 8 week I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA333118
VERSION AA333118.1 GI:1985362
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS Adams,M.D., Kerkavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gockayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

```



```

/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: parathyroid gland; Vector: pT7p3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTTACCACTGAAGTGGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTTTT
TTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7p3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."
BASE COUNT      95 a      36 c      33 g      93 t
ORIGIN

Query Match      4.6%; Score 240; DB 12; Length 257;
Best Local Similarity 100.0%; Pred. No. 3.1e-94;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5011 ggacaaccatgcaaatgtgaatagctggaagttcttggtattctaccacctcagta 5070
      |||
Db 257  GGACAACCACTGCAAAATGTGAAATAGTCTCTGGAATGTTCTTGGATTACTTTACACCTCAGTA 198
      |||

QY 5071 ttgattgtccagaattttctggccttcctggaatggaattttaaagaagaagatt 5130
      |||
Db 197  TTGATTGTCCGAGAAATTTCTGGCCCTTCATGGCAATGAAATTTTAAAGAAGAAGATT 138
      |||

QY 5131 taaagtattttaaagagtggtgtataaaataatgactgaattctttatcccat 5190
      |||
Db 137  TAAAGATATTTTAAAGAGTGTGTTATAAATAATGACTGAATTCCTTTATCCCAT 78
      |||

QY 5191 ttatcatcttccagtttttattatctactgtatcaataaaattctgtaattggaatg 5250
      |||
Db 77   TTATCATCTCTTCAGTTTTTATTATTAATCTACTGCTATCAATAAATTTCTGTAATTGAATG 18
      |||

RESULT 36
BE551003
LOCUS      BE551003      535 bp      mRNA      EST      10-AUG-2000
DEFINITION 7b67b03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3233261 3', mRNA sequence.
ACCESSION  BE551003
VERSION    BE551003.1 GI:9792695
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS    1 (bases 1 to 535)
TITLE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL    Tumor Gene Index
COMMENT    Unpublished (1997)
           Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
           Seq primer: -40UP from Gibco
           High quality sequence stop: 475.
           Location/Qualifiers

```

```

1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3233261"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT7p3D-Pac (Pharmacia) with a modified polylinker; plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      183 a      90 c      108 g      153 t
ORIGIN

Query Match      4.5%; Score 237; DB 137; Length 535;
Best Local Similarity 100.0%; Pred. No. 6.2e-93;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2919 aaggagagagagctcactaggaatgtttggtgaaaaataataatgaagcgaggag 2978
      |||
Db 167  AAGGAGAGAGAGCTCATGCTAGGCAATGTTGGTGAATAATAATGTAAAGCGGGAG 226
      |||

QY 2979 tatctgaacgacgacgtgtagtgaataattgtctcttaccagagatgtt 3038
      |||
Db 227  TATCTGAACGACGATGCTGTTAGTGAATAATTTATGTCCTCTTACCAGAGTAGTT 286
      |||

QY 3039 gttccatatcaattacccttttggcacatgacccagattatgtcaaaagtacaggatatt 3098
      |||
Db 287  GTTCCATATACAAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATT 346
      |||

QY 3099 gaacaactaaagatgtttaaagaatctcttggttcttggaataataatgaatgctt 3155
      |||
Db 347  GAACAACCTTAAAGATGTTAAAGAAATCTCTTGGTTTGTCTGGAATAATTTAATGGCT 403
      |||

RESULT 37
H93424
LOCUS      H93424      446 bp      mRNA      EST      01-DEC-1995
DEFINITION ys7a11.s1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:220700 3', mRNA sequence.
ACCESSION  H93424
VERSION    H93424.1 GI:1099752
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS    1 (bases 1 to 446)
           Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Narra, M., Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
           The WashU-Merck EST Project
           Unpublished (1995)
           Contact: Willson RK
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@wustl.edu
           High quality sequence stops: 239
           Source: IMAGE Consortium, LLNL
           This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
           Insert Length: 1379 Std Error: 0.00
           Seq primer: Promega -2ml3
           High quality sequence stop: 239.

```


1414920-1417991 and 1520904-1522439}. Subtraction by Bento Soares and M. Fatima Bonaldo.
BASE COUNT 92 a 59 c 71 g 135 t
ORIGIN

Query Match 5.3%; Score 278; DB 164; Length 357;
Best Local Similarity 99.7%; Pred. No. 6.4e-111;
Matches 328; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4707 ttatctgcggttaactaacagatcccgactgtttacttgatctattgattgaaag 4766
Db 29 TTATCTCGGCTTTAAACTGACAGTACCGACTGTTTATTGGATCTATTGATTGAAAG 88
QY 4767 aatttgtagatagattcaagcagtaactctgcaagtgtttgattgtattttctgca 4826
Db 89 AATTGTTAGATAGATCTTAAGCAGTAATCTGCACTGTTTGTATTGTTTCTGCA 148
QY 4827 attttactgtgaaaaaaattgttttcaacaattggtcatttcttgatgcaactat 4886
Db 149 ATTTTACTGTGAAAAAATTTGTTTCAACAATTGGTCACTTTCTTGATGTCATAT 208
QY 4887 ttgttgagagtttaaatggtctctccctttgtgtatctctacctagtttactcctggg 4946
Db 209 TTGTTGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTTACCTAGTGTTTACTCTGGG 268
QY 4947 cacccttaattcagagggtcaaaattgtctccattacacagaaggagcctctgat 5006
Db 269 CACCTTAATCTCAGAGGTGCTAAATGTCTGCCATTACACGAAGGATGCCCTCGAT 328
QY 5007 aggaggcaaccatgcaaaattgtgaata 5035
Db 329 AGGAGCACACCATGCAAAATGTGAATA 357

RESULT 34
BG387646
LOCUS 602412496F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4521095 5',
DEFINITION mRNA sequence.
ACCESSION BG387646
VERSION BG387646.1 GI:13281092
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 802)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: AFCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10419 row: p column: 24
High quality sequence stop: 637.
Location/Qualifiers
1. .802
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4521095"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 2.5 kb. Library enriched for

FEATURES
source

full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library.
BASE COUNT 270 a 149 c 167 g 216 t
ORIGIN

Query Match 5.1%; Score 269; DB 153; Length 802;
Best Local Similarity 99.7%; Pred. No. 5.6e-107;
Matches 389; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1158 gctattagacatgatgttattgtgtcaatagttacagctgtctaaaggatattctcttg 1217
Db 1 GCTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGTCTAAAGGATATTCTTCTG 60
QY 1218 gtcaatgatcacttaacttaattttgtgagagagagacattagacaaacgagtgagagta 1277
Db 61 GTCAATGATCATTACTTACTTAAATTTTGTGAGAGAGAACATTAGACAAACGATGGAGTA 120
QY 1278 cgcaaaagaagccatgatgggactgtcccaaatattataagaaatgctttcacagtcaag-c 1336
Db 121 CGCAAAAGAACCATGATGGGACTTGCCCAATTTATAGAAATATGCTTTACAGTCAGAC 180
QY 1337 agctggaaaagatgctgcaaaacagatagcatgatcaaaagacaaattgctacatatata 1396
Db 181 AGCTGGAAGAAGATGCTGCAAAACAGATAGCATGATGATCAAAAGACAAATTCGTACATATA 240
QY 1397 ttatcaaaatagattatgatgactgactactgttgaacggatctttgctcaatacattgt 1456
Db 241 TTATCAAAATAGTATGATGATCGACTACTTGTGTAACGGATCTTGTCTCAATACATGGT 300
QY 1457 tcctcacaatttagaaactacagacgagtgaaatgcttatattactgtatgcccact 1516
Db 301 TCCTCACAAATTTAGAACTACAGAACGGATGAATGCTTATATTACTTGTATGCCACACT 360
QY 1517 ggatttaaatgctgtgaagcattgaaatga 1546
Db 361 GGATTTAAATGCTGTGAAGCATTGAATGA 390

RESULT 35
AA854212/c

LOCUS a172a03.s1 Soares.parathyroid_tumor_NbHPA Homo sapiens cDNA clone
DEFINITION IMAGE:1401964 3', mRNA sequence.
ACCESSION AA854212
VERSION AA854212.1 GI:2941750
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 257)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo,
Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 554 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 182.
Location/Qualifiers
1. .257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1401964"


```

Db 301 ATTGAAAGAAATTGTTAGGATAGATCTTAAGCAGTAATCTGTCACTGTTGTAT 356
|||||
RESULT 30
AI140790/c
LOCUS      405 bp      mRNA      EST      13-FEB-1999
DEFINITION  ool5a04.x1 Soares.NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:1566222 3', mRNA sequence.
ACCESSION  AI140790
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 405)
AUTHORS   NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgaps-remail.nih.gov
          This clone is available royalty-free through LLNL ; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert Length: 544 Std Error: 0.00
          Seq primer: -40m13 fwd. ET from Amersham
          High quality sequence stop: 374.
FEATURES   Location/Qualifiers
            source
              1..405
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1566222"
                /clone_lib="Soares.NSF_F8_9W_OT_PA_P_S1"
                /lab_host="DH10B"
                /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                Equal amounts of plasmid DNA from five normalized
                libraries were mixed, and ss circles were made in vitro.
                Following HAP purification, this DNA was used as tracer in
                a subtractive hybridization reaction. The driver was
                PCR-amplified cDNAs from pools of 5,000 clones made from
                the same 5 libraries. The pools consisted of the following
                libraries and cloneIDs: Soares NBHSF pool 1:
                309384-310919, 323208-325895 Soares Nb2HP pool 1:
                145032-147335, 147720-148103, 148872-149255, 15002 -
                150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                758280-760583, 772104-774407 Soares NbHPA pool 1:
                304776-306311, 320136-322823, 326280-326663 Soares NBHOT
                pool 1: 723720-726407, 739080-740999 Subtraction by Bento
                Soares and M. Fatima Bonaldo."
            BASE COUNT  146 a 63 c 61 g 135 t
            ORIGIN
              1
              2
              3
              4
              5
              6
              7
              8
              9
              10
              11
              12
              13
              14
              15
              16
              17
              18
              19
              20
              21
              22
              23
              24
              25
              26
              27
              28
              29
              30
              31
              32
              33
              34
              35
              36
              37
              38
              39
              40
              41
              42
              43
              44
              45
              46
              47
              48
              49
              50
              51
              52
              53
              54
              55
              56
              57
              58
              59
              60
              61
              62
              63
              64
              65
              66
              67
              68
              69
              70
              71
              72
              73
              74
              75
              76
              77
              78
              79
              80
              81
              82
              83
              84
              85
              86
              87
              88
              89
              90
              91
              92
              93
              94
              95
              96
              97
              98
              99
              100
              101
              102
              103
              104
              105
              106
              107
              108
              109
              110
              111
              112
              113
              114
              115
              116
              117
              118
              119
              120
              121
              122
              123
              124
              125
              126
              127
              128
              129
              130
              131
              132
              133
              134
              135
              136
              137
              138
              139
              140
              141
              142
              143
              144
              145
              146
              147
              148
              149
              150
              151
              152
              153
              154
              155
              156
              157
              158
              159
              160
              161
              162
              163
              164
              165
              166
              167
              168
              169
              170
              171
              172
              173
              174
              175
              176
              177
              178
              179
              180
              181
              182
              183
              184
              185
              186
              187
              188
              189
              190
              191
              192
              193
              194
              195
              196
              197
              198
              199
              200
              201
              202
              203
              204
              205
              206
              207
              208
              209
              210
              211
              212
              213
              214
              215
              216
              217
              218
              219
              220
              221
              222
              223
              224
              225
              226
              227
              228
              229
              230
              231
              232
              233
              234
              235
              236
              237
              238
              239
              240
              241
              242
              243
              244
              245
              246
              247
              248
              249
              250
              251
              252
              253
              254
              255
              256
              257
              258
              259
              260
              261
              262
              263
              264
              265
              266
              267
              268
              269
              270
              271
              272
              273
              274
              275
              276
              277
              278
              279
              280
              281
              282
              283
              284
              285
              286
              287
              288
              289
              290
              291
              292
              293
              294
              295
              296
              297
              298
              299
              300
              301
              302
              303
              304
              305
              306
              307
              308
              309
              310
              311
              312
              313
              314
              315
              316
              317
              318
              319
              320
              321
              322
              323
              324
              325
              326
              327
              328
              329
              330
              331
              332
              333
              334
              335
              336
              337
              338
              339
              340
              341
              342
              343
              344
              345
              346
              347
              348
              349
              350
              351
              352
              353
              354
              355
              356
              357
              358
              359
              360
              361
              362
              363
              364
              365
              366
              367
              368
              369
              370
              371
              372
              373
              374
              375
              376
              377
              378
              379
              380
              381
              382
              383
              384
              385
              386
              387
              388
              389
              390
              391
              392
              393
              394
              395
              396
              397
              398
              399
              400
              401
              402
              403
              404
              405
              406
              407
              408
              409
              410
              411
              412
              413
              414
              415
              416
              417
              418
              419
              420
              421
              422
              423
              424
              425
              426
              427
              428
              429
              430
              431
              432
              433
              434
              435
              436
              437
              438
              439
              440
              441
              442
              443
              444
              445
              446
              447
              448
              449
              450
              451
              452
              453
              454
              455
              456
              457
              458
              459
              460
              461
              462
              463
              464
              465
              466
              467
              468
              469
              470
              471
              472
              473
              474
              475
              476
              477
              478
              479
              480
              481
              482
              483
              484
              485
              486
              487
              488
              489
              490
              491
              492
              493
              494
              495
              496
              497
              498
              499
              500
              501
              502
              503
              504
              505
              506
              507
              508
              509
              510
              511
              512
              513
              514
              515
              516
              517
              518
              519
              520
              521
              522
              523
              524
              525
              526
              527
              528
              529
              530
              531
              532
              533
              534
              535
              536
              537
              538
              539
              540
              541
              542
              543
              544
              545
              546
              547
              548
              549
              550
              551
              552
              553
              554
              555
              556
              557
              558
              559
              560
              561
              562
              563
              564
              565
              566
              567
              568
              569
              570
              571
              572
              573
              574
              575
              576
              577
              578
              579
              580
              581
              582
              583
              584
              585
              586
              587
              588
              589
              590
              591
              592
              593
              594
              595
              596
              597
              598
              599
              600
              601
              602
              603
              604
              605
              606
              607
              608
              609
              610
              611
              612
              613
              614
              615
              616
              617
              618
              619
              620
              621
              622
              623
              624
              625
              626
              627
              628
              629
              630
              631
              632
              633
              634
              635
              636
              637
              638
              639
              640
              641
              642
              643
              644
              645
              646
              647
              648
              649
              650
              651
              652
              653
              654
              655
              656
              657
              658
              659
              660
              661
              662
              663
              664
              665
              666
              667
              668
              669
              670
              671
              672
              673
              674
              675
              676
              677
              678
              679
              680
              681
              682
              683
              684
              685
              686
              687
              688
              689
              690
              691
              692
              693
              694
              695
              696
              697
              698
              699
              700
              701
              702
              703
              704
              705
              706
              707
              708
              709
              710
              711
              712
              713
              714
              715
              716
              717
              718
              719
              720
              721
              722
              723
              724
              725
              726
              727
              728
              729
              730
              731
              732
              733
              734
              735
              736
              737
              738
              739
              740
              741
              742
              743
              744
              745
              746
              747
              748
              749
              750
              751
              752
              753
              754
              755
              756
              757
              758
              759
              760
              761
              762
              763
              764
              765
              766
              767
              768
              769
              770
              771
              772
              773
              774
              775
              776
              777
              778
              779
              780
              781
              782
              783
              784
              785
              786
              787
              788
              789
              790
              791
              792
              793
              794
              795
              796
              797
              798
              799
              800
              801
              802
              803
              804
              805
              806
              807
              808
              809
              810
              811
              812
              813
              814
              815
              816
              817
              818
              819
              820
              821
              822
              823
              824
              825
              826
              827
              828
              829
              830
              831
              832
              833
              834
              835
              836
              837
              838
              839
              840
              841
              842
              843
              844
              845
              846
              847
              848
              849
              850
              851
              852
              853
              854
              855
              856
              857
              858
              859
              860
              861
              862
              863
              864
              865
              866
              867
              868
              869
              870
              871
              872
              873
              874
              875
              876
              877
              878
              879
              880
              881
              882
              883
              884
              885
              886
              887
              888
              889
              890
              891
              892
              893
              894
              895
              896
              897
              898
              899
              900
              901
              902
              903
              904
              905
              906
              907
              908
              909
              910
              911
              912
              913
              914
              915
              916
              917
              918
              919
              920
              921
              922
              923
              924
              925
              926
              927
              928
              929
              930
              931
              932
              933
              934
              935
              936
              937
              938
              939
              940
              941
              942
              943
              944
              945
              946
              947
              948
              949
              950
              951
              952
              953
              954
              955
              956
              957
              958
              959
              960
              961
              962
              963
              964
              965
              966
              967
              968
              969
              970
              971
              972
              973
              974
              975
              976
              977
              978
              979
              980
              981
              982
              983
              984
              985
              986
              987
              988
              989
              990
              991
              992
              993
              994
              995
              996
              997
              998
              999
              1000
              1001
              1002
              1003
              1004
              1005
              1006
              1007
              1008
              1009
              1010
              1011
              1012
              1013
              1014
              1015
              1016
              1017
              1018
              1019
              1020
              1021
              1022
              1023
              1024
              1025
              1026
              1027
              1028
              1029
              1030
              1031
              1032
              1033
              1034
              1035
              1036
              1037
              1038
              1039
              1040
              1041
              1042
              1043
              1044
              1045
              1046
              1047
              1048
              1049
              1050
              1051
              1052
              1053
              1054
              1055
              1056
              1057
              1058
              1059
              1060
              1061
              1062
              1063
              1064
              1065
              1066
              1067
              1068
              1069
              1070
              1071
              1072
              1073
              1074
              1075
              1076
              1077
              1078
              1079
              1080
              1081
              1082
              1083
              1084
              1085
              1086
              1087
              1088
              1089
              1090
              1091
              1092
              1093
              1094
              1095
              1096
              1097
              1098
              1099
              1100
              1101
              1102
              1103
              1104
              1105
              1106
              1107
              1108
              1109
              1110
              1111
              1112
              1113
              1114
              1115
              1116
              1117
              1118
              1119
              1120
              1121
              1122
              1123
              1124
              1125
              1126
              1127
              1128
              1129
              1130
              1131
              1132
              1133
              1134
              1135
              1136
              1137
              1138
              1139
              1140
              1141
              1142
              1143
              1144
              1145
              1146
              1147
              1148
              1149
              1150
              1151
              1152
              1153
              1154
              1155
              1156
              1157
              1158
              1159
              1160
              1161
              1162
              1163
              1164
              1165
              1166
              1167
              1168
              1169
              1170
              1171
              1172
              1173
              1174
              1175
              1176
              1177
              1178
              1179
              1180
              1181
              1182
              1183
              1184
              1185
              1186
              1187
              1188
              1189
              1190
              1191
              1192
              1193
              1194
              1195
              1196
              1197
              1198
              1199
              1200
              1201
              1202
              1203
              1204
              1205
              1206
              1207
              1208
              1209
              1210
              1211
              1212
              1213
              1214
              1215
              1216
              1217
              1218
              1219
              1220
              1221
              1222
              1223
              1224
              1225
              1226
              1227
              1228
              1229
              1230
              1231
              1232
              1233
              1234
              1235
              1236
              1237
              1238
              1239
              1240
              1241
              1242
              1243
              1244
              1245
              1246
              1247
              1248
              1249
              1250
              1251
              1252
              1253
              1254
              1255
              1256
              1257
              1258
              1259
              1260
              1261
              1262
              1263
              1264
              1265
              1266
              1267
              1268
              1269
              1270
              1271
              1272
              1273
              1274
              1275
              1276
              1277
              1278
              1279
              1280
              1281
              1282
              1283
              1284
              1285
              1286
              1287
              1288
              1289
              1290
              1291
              1292
              1293
              1294
              1295
              1296
              1297
              1298
              1299
              1300
              1301
              1302
              1303
              1304
              1305
              1306
              1307
              1308
              1309
              1310
              1311
              1312
              1313
              1314
              1315
              1316
              1317
              1318
              1319
              1320
              1321
              1322
              1323
              1324
              1325
              1326
              1327
              1328
              1329
              1330
              1331
              1332
              1333
              1334
              1335
              1336
              1337
              1338
              1339
              1340
              1341
              1342
              1343
              1344
              1345
              1346
              1347
              1348
              1349
              1350
              1351
              1352
              1353
              1354
              1355
              1356
              1357
              1358
              1359
              1360
              1361
              1362
              1363
              1364
              1365
              1366
              1367
              1368
              1369
              1370
              1371
              1372
              1373
              1374
              1375
              1376
              1377
              1378
              1379
              1380
              1381
              1382
              1383
              1384
              1385
              1386
              1387
              1388
              1389
              1390
              1391
              1392
              1393
              1394
              1395
              1396
              1397
              1398
              1399
              1400
              1401
              1402
              1403
              1404
              1405
              1406
              1407
              1408
              1409
              1410
              1411
              1412
              1413
              1414
              1415
              1416
              1417
              1418
              1419
              1420
              1421
              1422
              1423
              1424
              1425
              1426
              1427
              1428
              1429
              1430
              1431
              1432
              1433
              1434
              1435
              1436
              1437
              1438
              1439
              1440
              1441
              1442
              1443
              1444
              1445
              1446
              1447
              1448
              1449
              1450
              1451
              1452
              1453
              1454
              1455
              1456
              1457
              1458
              1459
              1460
              1461
              1462
              1463
              1464
              1465
              1466
              1467
              1468
              1469
              1470
              1471
              1472
              1473
              1474
              1475
              1476
              1477
              1478
              1479
              1480
              1481
              1482
              1483
              1484
              1485
              1486
              1487
              1488
              1489
              1490
              1491
              1492
              1493
              1494
              1495
              1496
              1497
              1498
              1499
              1500
              1501
              1502
              1503
              1504
              1505
              1506
              1507
              1508
              1509
              1510
              1511
              1512
              1513
              1514
              1515
              1516
              1517
              1518
              1519
              1520
              1521
              1522
              1523
              1524
              1525
              1526
              1527
              1528
              1529
              1530
              1531
              1532
              1533
              1534
              1535
              1536
              1537
              1538
              1539
              1540
              1541
              1542
              1543
              1544
              1545
              1546
              1547
              1548
              1549
              1550
              1551
              1552
              1553
              1554
              1555
              1556
              1557
              1558
              1559
              1560
              1561
              1562
              1563
              1564
              1565
              1566
              1567
              1568
              1569
              1570
              1571
              1572
              1573
              1574
              1575
              1576
              1577
              1578
              1579
              1580
              1581
              1582
              1583
              1584
              1585
              1586
              1587
              1588
              1589
              1590
              1591
              1592
              1593
              1594
              1595
              1596
              1597
              1598
              1599
              1600
              1601
              1602
              1603
              1604
              1605
              1606
              1607
              1608
              1609
              1610
              1611
              1612
              1613
              1614
              1615
              1616
              1617
              1618
              1619
              1620
              1621
              1622
              1623
              1624
              1625
              1626
              1627
              1628
              1629
              1630
              1631
              1632
              1633
              1634
              1635
              1636
              1637
              1638
              1639
              1640
              1641
              1642
              1643
              1644
              1645
              1646
              1647
              1648
              1649
              1650
              1651
              1652
              1653
              1654
              1655
              1656
              1657
              1658
              1659
              1660
              1661
              1662
              1663
              1664
              1665
              1666
              1667
              1668
              1669
              1670
              1671
              1672
              1673
              1674
              1675
              1676
              1677
              1678
              1679
              1680
              1681
              1682
              1683
              1684
              1685
              1686
              1687
              1688
              1689
              1690
              1691
              1692
              1693
              1694
              1695
              1696
              1697
              1698
              1699
              1700
              1701
              1702
              1703
              1704
              1705
              1706
              1707
              1708
              1709
              1710
              1711
              1712
              1713
              1714
              1715
              1716
              1717
              1718
              1719
              1720
              1721
              1722
              1723
              1724
              1725
              1726
              1727
              1728
              1729
              1730
              1731
              1732
              1733
              1734
              1735
              1736
              1737
              1738
              1739
              1740
              1741
              1742
              1743
              1744
              1745
              1746
              1747
              1748
              1749
              1750
              1751
              1752
              1753
              1754
              1755
              1756
              1757
              1758
              1759
              1760
              1761
              1762
              1763
              1764
              1765
              1766
              1767
              1768
              1769
              1770
              1771
              1772
              1773
              1774
              
```


REFERENCE 1 (bases 1 to 410)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert Length: 719 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 398.

FEATURES

source
 1..410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1603350"
 /clone_lib="NCI CGAP Lu5"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 neuroendocrine lung carcinoid, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."
 112 a 59 c 88 g 151 t

BASE COUNT

ORIGIN
 Query Match 5.9%; Score 311; DB 14; Length 410;
 Best Local Similarity 99.7%; Pred. No. 2.1e-125;
 Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4458 ggtcaagcttgaggctgaataaagcctttgatgcacaaaatgggactgctgaagatgga 4517
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ggtcaagcttgaggctgaataaagcctttgatgcacaaaatgggactgctgaagatgga 60
 QY 4518 cagttggaccttactttggtgacccacatactttggtcacatgctttagccatacaca 4577
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 cagttggaccttactttggtgacccacatactttggtcacatgctttagccatacaca 120
 QY 4578 tggtaacattgactatggagctcttgaaagtgtaatgtgcgactggtatgtagacataa 4637
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 tggtaacattgactatggagctcttgaaagtgtaatgtgcgactggtatgtagacataa 180
 QY 4638 aagaagaacttgaataatcttttttttttttttttttttttttttttttttttttttt 4697
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 aagaagaacttgaataatcttttttttttttttttttttttttttttttttttttttt 240
 QY 4698 gtatagctttttatctgcggctttaaactgacagctaccacactgtttattggatctattga 4757
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 gtatagctttttatctgcggctttaaactgacagctaccacactgtttattggatctattga 300
 QY 4758 ttgtgaaagaattgttaggatatagatctttaagcagtaatactgcagtggtttgtatttga 4817
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 ttgtgaaagaattgttaggatatagatctttaagcagtaatactgcagtggtttgtatttga 360
 QY 4818 tt 4819
 ||
 Db 361 tt 362

RESULT 29

AA911125

LOCUS AA911125 356 bp mRNA EST 21-APR-1998
 DEFINITION oe69a12.s1 NCI-CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1416862 3',
 mRNA sequence.
 ACCESSION AA911125
 VERSION AA911125.1 GI:3050415
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 356)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert Length: 725 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 328.

FEATURES

source
 1..356
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1416862"
 /clone_lib="NCI CGAP Lu5"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 neuroendocrine lung carcinoid, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."
 96 a 52 c 80 g 128 t

Query Match 5.8%; Score 305; DB 13; Length 356;
 Best Local Similarity 99.7%; Pred. No. 9.2e-123;
 Matches 355; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4457 tggtaacattgactatggagctcttgaaagtgtaatgtgcgactggtatgtagacata 4516
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 tggtaacattgactatggagctcttgaaagtgtaatgtgcgactggtatgtagacata 60
 QY 4517 acagttggaccttactttggtgacccacatactttggtcacatgcttttagccatacac 4576
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 acagttggaccttactttggtgacccacatactttggtcacatgcttttagccatacac 120
 QY 4577 atcggttaacattgactatggagctcttgaaagtgtaatgtgcgactggtatgtagacata 4636
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 atcggttaacattgactatggagctcttgaaagtgtaatgtgcgactggtatgtagacata 180
 QY 4637 aagaagaacttgaataatcttttttttttttttttttttttttttttttttttttttt 4696
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 aagaagaacttgaataatcttttttttttttttttttttttttttttttttttttttt 240
 QY 4697 tgtatagctttttatctgcggctttaaactgacagctaccacactgtttattggatctattg 4756
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 tgtatagctttttatctgcggctttaaactgacagctaccacactgtttattggatctattg 300
 QY 4757 atttgaagaagaattttgttaggatatagatctttaagcagtaatactgcagtggtttgtattg 4812


```

/db_xref="taxon:9606"
/clone="IMAGE:460933"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1: Sfil (ggccattggcc); Site_2: Sfil (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT      225 a  129 c  144 g  231 t
ORIGIN

```

```

Query Match      5.9%; Score 313; DB 153; Length 729;
Best Local Similarity 99.7%; Pred. No. 2.8e-126;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 470 tgcctgggtcaagctatataacatagcttgggttagagtagaagaattttttac 529
|
|
|
Db 42 TGCTGGGTCAAGTCAATATACATATGCTTTGAGTTAGAGATAGCAATGAAATTTTCAC 101
|
|
|
QY 530 ccagctatacagaacaccttttcagttataaacaatggccacaatcagaagctccat 589
|
|
|
Db 102 CCAGCTATACAGAACCTTATTTTCAGTTATTAACAATGGCCCAATCAGAAAGTCCATAT 161
|
|
|
QY 590 gcacatggtagagacctatgagctctattttgtgaaggtgatacagtgctcaggagct 649
|
|
|
Db 162 GCACATGGTAGACCTATGAGCTCTATTATTGTGAAGGTGATACAGTGTCTCAGGAGCT 221
|
|
|
QY 650 ttggatacaggttttagtaaatctgtacctgtctataagaatttaacaagaacgata 709
|
|
|
Db 222 TTTGGTACAGGTTTATAGTAATCTGTGTACCTGTCTCAAGAATTTAAACAAGCAGATA 281
|
|
|
QY 710 tgattggcaaggcttactgaagagacagctcaagctattgagccatattaccac 769
|
|
|
Db 282 TGATTGGCAAGGCTTACTGAAGAGCAGCTCAAGCTATTGTGAGCCATATATTACCAA 341
|
|
|
QY 770 ttttttaacaggttctgatgttgggaacaacatctatcagcgattttgtcagagcattg 829
|
|
|
Db 342 TTTTAAATCAGGTTCTGATGCTTGGGAAACATCTATCAGCGATTGTGTCAGAGCATGT 401
|
|
|
QY 830 cttt 833
|
|
|
Db 402 CTTT 405

```

```

RESULT 27
AI299646      406 bp  mRNA      EST      29-JAN-1999
LOCUS      gn12a10.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1898010 3',
DEFINITION      mRNA sequence.
ACCESSION      AI299646
VERSION      AI299646.1 GI:3959196
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 406)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.

```

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/obrp/image/image.html
Insert length: 548 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 394.

FEATURES
Location/Qualifiers
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1898010"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 112 a 58 c 87 g 149 t
ORIGIN

Query Match 5.9%; Score 311; DB 18; Length 406;
Best Local Similarity 99.7%; Pred. No. 2.1e-125;
Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 4458 ggtcaagcttgagctgaataaagccttgcacacaaaatggagactgctgaagagtgga 4517
|
|
|
Db 1 GGTCACAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAGAGTGGA 60
|
|
|
QY 4518 cagtgtgaccttacttgggtgaccccatcacatttgggtcacatgctttagccatcacaca 4577
|
|
|
Db 61 CAGTTGGACCTTACTTTGGTGACCCCATACGTTTGTGTCACATGCTTTAGCCATACACA 120
|
|
|
QY 4578 tggtaacattgactatggagctcttggtaagtgtaattgtgcgattggtatgacataaa 4637
|
|
|
Db 121 TGTTAACATTGACTATGAGTCTTGTGAAAAGTGTAAATGTGCGATGGCTATGTAGACATAA 180
|
|
|
QY 4638 agaagaaactgttaaatatcttttcttttttttttttttttttttttttttttttttt 4697
|
|
|
Db 181 AGAAGAAACCTGTAAATATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 240
|
|
|
QY 4698 gtatagcttttattctgcggctttaaactgacagtcaccgactgtttattggtatttga 4757
|
|
|
Db 241 GTATAGCTTTTATCTGCGGCTTTAACTGACAGTACCGACTGTTTATGATCTATGTA 300
|
|
|
QY 4758 ttgaaaagaatttggtagtagatcttaagcagtagtaaatctgctcagtgctttgtatttga 4817
|
|
|
Db 301 TTTGAAAAGAATTGTTAGGATAGATCTTAAGCAGTAGTAATCTGTGCTGCTTTGTTATTGTA 360
|
|
|
QY 4818 tt 4819
|
|
|
Db 361 TT 362

```

```

RESULT 28
AA987361
LOCUS      or92904.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1603350 3',
DEFINITION      mRNA sequence.
ACCESSION      AA987361
VERSION      AA987361.1 GI:3172725
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```


/clone="CuAAJC12"			
/clone_lib="Cu"			
/tissue_type="adrenal cortico adenoma for Cushing's syndrome"			
/dev_stage="Adult"			
/note="Vector: pBluescript sk(-)"			
BASE COUNT	221 a 113 c 150 g 253 t	1	others
ORIGIN			
Query Match	6.1%	Score 324;	DB 32; Length 738;
Best Local Similarity	99.7%	Pred. NO.4.2e-131;	
Matches 374;	Conservative	0;	Mismatches 1; Indels 0; Gaps
QY	4	gagagcccgagtgagcgagtagcgagtcgcgaacccgaggggtagaataattctctg	6
DB	2	GAGAGCCCGGAGTCAGCGGAGTAGCGAGTCGGCAACCCGAGGGGTAGAAATATTTCTG	6
QY	64	tcaatgctcatcaagactagaccatgatggaaaaattacatcatccctcgctggggtc	1
DB	62	TCATGCTCATTCAAAGACTAGACCAATGATGGAAAAATACATATCCGCTGGGGTCA	1
QY	124	aggaataatcagataaaatatactaaaggagagatggtgagacgattaaagatggtgtga	1
DB	122	AGAAATATCAGATAAAATATCTAAAGAGGAGATGCTGAGACGATTAAAGATGGTTGTGA	1
QY	184	aaacttttatgatgacacgaggaactctgaagaagaagagagctttatttaaacctag	2
DB	182	AACTTTTATGATATGGACAGGACTCTGAAGAAGAAGAGGCTTTATTATTAACCTAG	2
QY	244	cttacatcttcttcagatctttttctcaagcatctcgtgtaaagatgttcgcttactgg	3
DB	242	CTTTACATCTTGCTTCAGATTTTTTTCTCAAGCATCTCTGATAAAGATGTCGTTACTGG	3
QY	304	tagctctgctcttgctgatattttcagagattatgcttcctgaagctcttaacatccc	3
DB	302	TAGCCTGCTGCCCCGTGATATTTTCAGAGATTATGCTCCTGAAGCTCCTTACACATCCC	3
QY	364	ctgataaacctaaagg 378	
DB	362	CTGATAAACTAAAGG 376	
RESULT 26			
BG432960			
LOCUS	602495783F1 NTH_MGC_75	Homo sapiens cDNA clone IMAGE:4609533	14-MAR-2000
DEFINITION	mRNA sequence.		
ACCESSION	BG432960		
VERSION	BG432960.1	GI:13339466	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 729)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLM1353 row: e column: 22 High quality sequence stop: 722. Location/Qualifiers 1..729 /organism="Homo sapiens"		
FEATURES			
source			


```

DEFINITION      7n43b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3567291 3',
mRNA sequence.
ACCESSION       BF111072
VERSION         BF111072.1 GI:10940762
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 525)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE           Tumor Gene Index
JOURNAL         Unpublished (1997)
COMMENT         Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs@remail.nih.gov
                  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
                  Emmert-Buck, M.D., Ph.D.
                  cDNA Library Preparation: M. Bento Soares, Ph.D.
                  DNA Sequencing by: Greg Lennon, Ph.D.
                  Clone distribution: NCI-CGAP clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL, send email to:
                  info@image.llnl.gov
                  Seq primer: -40UP from Gibco
                  High quality sequence stop: 470.
FEATURES        Location/Qualifiers
source          1..525
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3567291"
                /clone_lib="NCI_CGAP_Lu24"
                /tissue_type="carcinoid"
                /lab_host="DH10B"
                /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                modified polylinker; Plasmid DNA from the normalized
                library NCI_CGAP_Lu5 was prepared, and ss circles were
                made in vitro. Following HAP purification, this DNA was
                used as tracer in a subtractive hybridization reaction.
                The driver was PCR-amplified cDNAs from a pool of 5,000
                clones made from the same library (cloneIDs
                1414920-1417991 and 1520904-1522439). Subtraction by Bento
                Soares and M. Fatima Bonaldo. "
BASE COUNT      133 a      81 c      109 g      201 t
ORIGIN
Query Match      7.2%; Score 381; DB 144; Length 525;
Best Local Similarity 99.6%; Pred. No. 4.1e-156;
Matches 481; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4453 ttttggctgaagcttgaggctgaataaagccttttgatgcacaaaatgggactgctgaaga 4512
Db 1 TTTTGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAGA 60

QY 4513 gtgacagtgacaccttaattggtgacccatacatattggtgacatgcttagccat 4572
Db 61 GTGACAGTGGACCTTACTTTGGTGACCCCATACGCTTGGTGTCACATGCTTTAGCCAT 120

QY 4573 acacatggtlaacatgaactagatgagctcttgaaagttaagtgcgatggctataga 4632
Db 121 ACACATGGTAAACATTGACATGAGGCTTGTGAAAGTGAATGTCGATGGCTATGTAGA 180

QY 4633 cataaagaagaactgtgaaatcatcttttcttttttttaattgattctgattctgaag 4692
Db 181 CATAAAGAAGAACTGTAAATATCTTTTCTTTTAAATGTTCTGATTTCTGGAAG 240

QY 4693 tgcctgtatagcttttatctgcggctttaaactgacagtcaccgactgttatgatatc 4752
Db 241 TGCCTGTATAGCTTTTATCTGCGGCTTTAACTGACAGTACCCGACTGTTTATTGGATCT 300

QY 4753 attgatttgaagaactttgttaggatacatcttaagcagtaactctgcagttgttat 4812
|||||

```

```

Db 301 ATTGATTGAAAAGAAATTGTTAGGATAGATCTTAAGCAGTAATCTGTCACTGTTGTAT 360
QY 4813 ttgtatttctgcaatttacttgtaaaaaaatttggtttttcaacaattggtgtcatttt 4872
|||||
Db 361 TTGTATTCTCTGCAATTTACTGTGAAAAAATTTGTTTCAACAATTTGCTGTCATTT 420
QY 4873 cttgactgctacatttctgtggagagtaaatggtctctcttccctttgtgtatcttaacctag 4932
|||||
Db 421 CTTGATGTCACTATTATTGTTGGAGAGTAAATGCTCTCTCCCTTTGTGTATCTTACCTAG 480
QY 4933 tgt 4935
Db 481 TGT 483

RESULT 22
BF446410
LOCUS          BF446410      390 bp      mRNA      EST      01-DEC-2000
DEFINITION    7p38h09.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3648232 3',
mRNA sequence.
ACCESSION     BF446410
VERSION       BF446410.1 GI:11511548
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 390)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE         Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
                  Email: cgapbs@remail.nih.gov
                  Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
                  Emmert-Buck, M.D., Ph.D.
                  cDNA Library Preparation: M. Bento Soares, Ph.D.
                  DNA Sequencing by: Greg Lennon, Ph.D.
                  Clone distribution: NCI-CGAP clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL, send email to:
                  info@image.llnl.gov
                  Seq primer: -40UP from Gibco.
FEATURES      Location/Qualifiers
source        1..390
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3648232"
              /clone_lib="NCI_CGAP_Pr28"
              /sex="male"
              /dev_stage="adult"
              /lab_host="DH10B"
              /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
              with a modified polylinker; Plasmid DNA from the
              normalized library NCI_CGAP_Pr22 was prepared, and ss
              circles were made in vitro. Following HAP purification,
              this DNA was used as tracer in a subtractive hybridization
              reaction. The driver was PCR-amplified cDNAs from a pool
              of 5,000 clones made from the same library (cloneIDs
              985608-986759, 1101192-1101959, and 1217928-1220615).
              Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT    105 a      56 c      83 g      146 t
ORIGIN
Query Match      7.0%; Score 368; DB 148; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.1e-150;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4452 tttttggctcaagcttgaggctgaataaagccttttgatgcacaaaatgggactgctgaag 4511
|||||
Db 1 TTTTGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAG 60

```


single-stranded DNA preparation of NCI_CGAP_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising a pool of clones from NCI_CGAP_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI_CGAP_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI_CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG_LIB=NCI_CGAP_Lu5

TAG_TISSUE=lung

TAG_SEQ=CAAC

BASE COUNT 167 a 79 c 96 g 148 t

Query Match 7.3%; Score 384; DB 149; Length 490;
Best Local Similarity 99.8%; Pred. No. 2e-157;
Matches 434; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2745 tgttaccatgaatcatcacattagaacaatcatcagctatgtcattagctatcaacgat 2804

Db 13 TGTATACCATGNAATCATCATATAGACATATACAGCTATGTGCTATGATCAACGAT 72

QY 2805 gaatgctatcaagtaagacaagtgttggccagagaacttcacaaagcccttcccgttta 2864

Db 73 GAATGCTATCAAGTAAGACAAGTGTGTGCCAGAACTTCACAAAGGCCTTTCCCGTTTA 132

QY 2865 cggcttccacttgatataaggcaatctgtgccttctgtgcacaaagatcctgttaaggag 2924

Db 133 CGGCTCCACTTGATATATGGCAATCTGTGCCCTTTGTGCAAAAGATCTGTATAAGGAG 192

QY 2925 agaagagctcatgtaggcaatgttggtagaaataataatgaagcgggagatctg 2984

Db 193 AGAGAGCTCATGCTAGGCAATGTTGGTGAAATAATATTAATGTAAAGCGGGATATCTG 252

QY 2985 aagcagcatgagctgttagtgaataattattgtctctctaccagagatgttgttcca 3044

Db 253 AAGCAGCATGCAGCTGTAGTGAATAATTATTGCTCTCTACCCAGAGATGTTGTTCCTCA 312

QY 3045 tatacaattcaccttttggcacatgacccagattatgtcaaaagtacaggatattgaacaa 3104

Db 313 TATAAATTACCTTTTGGCACATGACCCAGATGATGTCAAAGTACAGGATATTGAACAA 372

QY 3105 cttaaagatgttaagaatgtcttggttgttctggaataattaatggctaaataatgaa 3164

Db 373 CTAAAGATGTTAAAGATGTCCTTGGTTGTTCTGGAATATTAATTAATGGCTAAATAATGAA 432

QY 3165 aataacagtcacgct 3179

Db 433 AATAACAGTCACGCT 447

RESULT 20

AI275866

LOCUS

DEFINITION q168c06.x1 Soares_NHMPu_S1 Homo sapiens CDNA clone IMAGE:1877482

3', mRNA sequence.

ACCESSION AI275866

VERSION AI275866.1 GI:3898140

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 716 Std Error: 0.00
Seq primer: -40UP from Gibco

High quality sequence stop: 394.

FEATURES

source

1. 433
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1877482"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1; Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 117 a 62 c 90 g 163 t
ORIGIN

Query Match 7.2%; Score 381; DB 18; Length 433;
Best Local Similarity 99.8%; Pred. No. 4.1e-156;
Matches 431; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4458 ggtcaagcttgagctgaataaagcctttagtcacaaaatggagactgctgaagtgga 4517

Db 1 GGTCAAGCTTGAGCTGAATAAAGCCTTTGATGCACAAAATGGGACTGCTGAAGAGTGGA 60

QY 4518 cagtgacacttacttggtagcccccatacatttgggtcacatgctttagccatacaca 4577

Db 61 CAGTGGACCTTACTTTGGTGACCCCATACATTTGTGTACATGCTTTAGCCATACACA 120

QY 4578 tggtaacatgactatgagctcttggtaagagtgtaagtgcgctatgtatgacataa 4637

Db 121 TGGTAAACATTGACTATGAGCTCTTGTGAAAGTGTAAATGTGGATGGCTATGTAGACATAA 180

QY 4638 agaagaactgttaaatatcttttcttttttaatttttctgatttctggaagtgctt 4697

Db 181 AGAAGAACCTGTAAATATCTTTTTCTTTTTTAATGTTCTTGATTTCTGAAGTGCTT 240

QY 4698 gtatagcttttctgctgctttaaactgcagtcaccgactgtttattggtatctattga 4757

Db 241 GTATAGCTTTTATCTCGGCTTTAAACTGCAGTACCCGACTGTTTATGGATCTATTGA 300

QY 4758 ttgaaagaagaatttggtagatagatcttaagcagtgtaactctgcagtgcttatttgta 4817

Db 301 TTTGAAAGAAGAAATTTGTAGGATAGATCTTAAGCAGTAATCTGTGAGTGTGTTGTTGTA 360

QY 4818 ttttctcaattttactgtgaaaaaatttttttcaacaattggtgctcatttcttga 4877

Db 361 TTCTCTCAATTTTACTGTGTAAGAAAAAATTTGTTTTTCAACAATTTGGTGTCTATTTCCTTGA 420

QY 4878 tgtcactatttg 4889

Db 421 TGTCACTATTG 432

RESULT 21

BF111072

LOCUS

BF111072

525 bp

mRNA

EST

20-OCT-2000


```

Db 229 AAATAGTCTGAAGTCTTGGATTACTTTACACCTCAGTATGATTTGCCCCAGAAATTTT 170
QY 5091 ctggccttcctggaatgaaatcttaagaagaagatttaagatttttaatttttaa 5150
Db 169 CTGGCCTTTTCATGGCAATGAAATTTTAAAGAGAAAGATTTTAAAGATTTTAA 110
QY 5151 gaggctgtatataaataaagtactgaatctctttatccatctttatcatcctttcagtttt 5210
Db 109 GAGTGTGTTATATAAATAATGTACTGAATTTCTTTATCCCATTTTATCATCCTTTTCAGTTTT 50
QY 5211 tattaactactgtatcataataaaattctgtaatttgaatgag 5252
Db 49 TATTAATCTACTGTATCAATAAATAATCTGTAAATTTGAATGAG 8

RESULT 18
AI681422
LOCUS AI681422 486 bp mRNA EST 16-DEC-1999
DEFINITION tx46e01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272632 3',
mrna sequence.
ACCESSION AI681422
VERSION AI681422.1 GI:4891593
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 486)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 714 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 460.
FEATURES
Location/Qualifiers
1..486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2272632"
/tissue_type="carcinoid"
/lab_host="DH10B"
note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522435). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 125 a 73 c 101 g 187 t
ORIGIN
1..486
Query Match 7.3%; Score 384; DB 23; Length 486;
Best Local Similarity 99.6%; Pred. No. 2e-157;
Matches 484; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4457 tggctcagcttgagctgaataagcctttgacacaaatgggactgctgaagtggtg 4516
Db 1 TGGTCAAGCTTGAGGCTGAATAAAGCCTTTGATGCACAAATGGGACTGCTGAAGAGTGG 60

```

```

QY 4517 acagttggaccttaocttttggtgaccccatcacatttggtcacatgctcttagccatacac 4576
Db 61 ACAGTTGGACCTTACTTTGGTGACCCCATACATGTTTGGTGCACATGCTTTAGCCATACAC 120
QY 4577 atggttaacattgactgagctcttggtaagtgtaagtgcgcatgagcattgtagacata 4636
Db 121 ATGGTAACATTTGACTATGGAGTCTTTGTGAAAGTGTAAATGTGCGATGGCTATGTAGACATA 180
QY 4637 agagaagaactgttaaatctcttttctttttaaagtttctctgattctcgaagtgcct 4696
Db 181 AAGAAGAACTTGTAAATATCTTTTCTTTTAAAGTTTCTGATTTCTGAAGTGCT 240
QY 4697 tgtatagcttttctatctgcgctttaaactgacagtgacccgactggtttattggatctattg 4756
Db 241 TGTATAGCTTTTATCTCGCGCTTTAAACTGCACAGTACCCGACTGTTTATTGGATCTATTG 300
QY 4757 atttgaagaagaattttttaggtagatagattcctaagcagtaactctcagtggtttatttgt 4816
Db 301 ATTGTGAAAGAAATTTCTTAGGATAGATCTTAAGCAGTAATCTCTCAGTGTGTTGTTATTTGT 360
QY 4817 attttctgcaattttactgtgaaaaaaatttcttccaacaatttggtgctcatttcttg 4876
Db 361 ATTCTCTGCAATTTTACTGTGAAAAAAATTTGTTTCAACAATTTGGTGTCAATTTCTTG 420
QY 4877 atgtcactatttggtagagtgtaaatggtctctctcccttggcttcttaccctagtggtt 4936
Db 421 ATGTCACTATTGTTGGAGAGTAAATGGTCTCTCCCTTTGTGTATCTTACCTAGTGTT 480
QY 4937 tactcc 4942
Db 481 TACTCC 486

RESULT 19
BF511856
LOCUS BF511856 490 bp mRNA EST 06-DEC-2000
DEFINITION UI-H-B14-aps-d-12-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3088558 3', mRNA sequence.
ACCESSION BF511856
VERSION BF511856.1 GI:11595154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 490)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
Location/Qualifiers
1..490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3088558"
/lab_host="DH10B (Life Technologies)"
note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub8
is a subtracted library derived from NCI_CGAP_Sub5. The
NCI_CGAP_Sub8 library had 2.5 million recombinants. A

```


IMAGE:1627886 3', mRNA sequence.
 ACCESSION AT016896
 VERSION AT016896.1 GI:3231232
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 461)
 AUTHORS NCI-CCAG <http://www.ncbi.nlm.nih.gov/ncicag>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 Insert Length: 548 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 420.
 FEATURES
 Location/Qualifiers
 1..461
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1627886"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pTT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CCAG-SCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 182 a 76 c 72 g 131 t
 ORIGIN
 Query Match 7.7%; Score 406; DB 15; Length 461;
 Best Local Similarity 100.0%; Pred. No. 4.4e-167;
 Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4821 tctgcaatttactgtgaaaaaatttcttcaacaattggtgtcatttcttgatg 4880
 |||||||
 Db 436 TCTGCAATTTTACTGTGAAAAAATTTGTTTCAACAATGGTGTCATTTCCTTGATGT 377
 |||||||
 Qy 4881 cactatttggagagtaaatggtctctccctcttctgtatcttactagtgttact 4940
 |||||||
 Db 376 CACTATTTTGGAGAGTAATGGTGCTCTCCCTTTGTGTATCTTACTAGTGTACT 317
 |||||||
 Qy 4941 cctggacccttaattctcagagtgctaaattgtctgcatcacacagaagatgcc 5000
 |||||||
 Db 316 CCTGGCACCTTAACTCTTCAGAGTGCTAAATGTCTGCCATTACACCAAGAGATGCC 257
 |||||||
 Qy 5001 tctgtaggaggaacacacatgcaaatgtgaaatagtcctgaagtcttgattacttta 5060
 |||||||
 Db 256 TCTGATAGGAGGACACCAACCATGCAATTTGTGAATAGTCTCCTGAAGTCTTGACTTTA 197
 |||||||
 Qy 5061 cactcagattgattgtccagaatttctggccttctatgccaatgaaatttaag 5120
 |||||||
 Db 196 CACCTCAGTATGTATTTCTCCAGAAATTTCTGGCTTTTCATGGCAATGAAATTTTAAAG 137
 |||||||
 Qy 5121 aagaaagattaaagtatttttaatttaagagtggtgtataaaaaataagtactgaattc 5180
 |||||||
 Db 136 AAGAAGATTAAAGTATTTTAAAGAGTGTTTATAAATAATGATGATGAATTC 77
 |||||||
 Qy 5181 ttatccattttatcatccttccagtttttatttaataactactgtat 5226
 |||||||

Db 76 TTTATCCCAATTTTATCATCCCTTCAGTTTATTAATTAATCTACTGTAT 31
 |||||||
 RESULT 17
 LOCUS AA625961/c 409 bp mRNA EST 15-OCT-1997
 DEFINITION zu92e08.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:745478
 3', mRNA sequence.
 ACCESSION AA625961
 VERSION AA625961.1 GI:2538348
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 409)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WASHU-NCI human EST Project
 Unpublished (1997)
 COMMENT Contact: Wilton RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 378.
 FEATURES
 Location/Qualifiers
 1..409
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:745478"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pTT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 155 a 67 c 66 g 121 t
 ORIGIN
 Query Match 7.6%; Score 402; DB 9; Length 409;
 Best Local Similarity 100.0%; Pred. No. 2.5e-165;
 Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4851 tttaacaattggtgtcatttcttctgtatgtcactatttggagagtgtaaatgtctct 4910
 |||||||
 Db 409 TTTCACAATTTGGTGTCATTTCTTGATGTCTACATTTTGTGGAGAGTTAAATGGTCTCT 350
 |||||||
 Qy 4911 tcccttggatcttactagtgttttactcctgggacccttaattctcagagtgcta 4970
 |||||||
 Db 349 TCCCTTTGTATCTTACTAGTGTGTTTACTCTGGGCCCTTAACTTCAGAGTGCTA 290
 |||||||
 Qy 4971 aatgtctgccattacacagaagatgcctctatagagacaacatcaaatgtg 5030
 |||||||
 Db 289 AATTGTCGCCATTACACCAAGAGATGCCCTCTGTAGAGAGCAACCATGCAAAATGTG 230
 |||||||
 Qy 5031 aaatagtcctgaagtcttctgttacttctcacctcagttattgttccccagaatttt 5090
 |||||||

FEATURES

Location/Qualifiers
1. 578

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3008877"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 203 a 97 c 112 g 163 t 3 others

FEATURES

source

Query Match 8.0%; Score 422; DB 120; Length 578;
Best Local Similarity 99.8%; Pred. No. 4.2e-174;
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2755 aaatcatcattagacaataatcagctatgtgcatatgattatcaacgatgaatgctatc 2814

|||||

Db 4 AAATCATCATATTAGAACAAATATACATGTGTGTCATGTAGTATCAACGATGAATGCTATC 63

|||||

QY 2815 aagtaagacaagtgttgcacgaacttcacaaagcccttccgctttacggctccac 2874

|||||

Db 64 AGTAGACAAAGTGTGGCCGCAAACTTCACAAAGCCCTTCCCGTTTACGGCTCCAC 123

|||||

QY 2875 ttgagtatatgcaatctgtgccccttctgcaaaagatcctgtaaggagagaagatc 2934

|||||

Db 124 TTGAGTATATGCAATCTGCGCTTGTGCAAAAGATCCTGTAAAGGAGAGAGAGCTC 183

|||||

QY 2935 atgctaggcaatgttgggtgaaataataatgtaaggcggagtagtctggaagcagatg 2994

|||||

Db 184 ATGCTANGCAATGTTGGTGAATAATATAATGTAAGCGGGAGTATCTGAAGCAGCATG 243

|||||

QY 2995 cagctttagtgaataattatgtctctctaccagatgattgtgttccatatacaattc 3054

|||||

Db 244 CAGCTTTAGTGAAAATTTATGTCTCTCTACCAGATGATGTTGTTCCATATACAAATTC 303

|||||

QY 3055 acctttggcacatgaccagattatgtcaaaagtagcaggtatttgacaacattaaagatg 3114

|||||

Db 304 ACCTTTGGCACATGACCCAGATTTATGTCAAAGTACAGGATATTGAACAACCTTAAAGATG 363

|||||

QY 3115 ttaaagaattgttgggttctgtggaataattaatggcctaataatgaaataaacagtc 3174

|||||

Db 364 TTAAGAATGTCTTGGTGTGCTGGAATATTAATGCTTAAATGCTTAAATGAAATTAACAGTC 423

|||||

QY 3175 acgctttatcagaagatggttagaataatttaacaaacaaagaatgccaa 3227

|||||

Db 424 ACGCTTTATCAGAAGATGGTAGAAAATTTAAACAACAACAAGATGCCAA 476

|||||

RESULT 15

AA563884/c

LOCUS nk19e08.s1 NCI_CGAP_Coll Homo sapiens cDNA clone IMAGE:101390 3',

DEFINITION mRNA sequence.

ACCESSION AA563884

VERSION AA563884.1 GI:2335523

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

JOURNAL

COMMENT

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Elias Campo,

M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 868 Std Error: 0.00

Seq primer: -40ml3 fwd. Et from Amersham

High quality sequence stop: 365.

Location/Qualifiers

1. 469

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:101390"

/clone_lib="NCI_CGAP_Coll"

/tissue_type="tumor"

/lab_host="SOLR (kanamycin resistant)"

; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo

dt. Multiple colon tumors. 5' adaptor sequence: 5'

GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'

CTCAGATGTTTTTTTTTTT 3' Average insert size: 1.1 kb."

BASE COUNT 184 a 81 c 71 g 133 t

ORIGIN

Query Match 7.9%; Score 418; DB 9; Length 469;
Best Local Similarity 99.8%; Pred. No. 2.4e-172;
Matches 468; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4748 gatctattgattgaaagaattgttagtagatcttaagcagtaactgtcagtggt 4807

|||||

Db 469 GATCTATTGATTGAAAGAATTTGTAGGATAGATCTTAAAGCAGTAATCTGTCAGTGT 410

|||||

QY 4808 tgtattgtatcttctgcaatttactgtgaaaaaaattgttttcaacaattgtgtc 4867

|||||

Db 409 TGTATTTGTTCTGTGCAATTTTACTGTGAAAAAAATTTGTTTCAACAATTTGGTGTGTC 350

|||||

QY 4868 attttcttgatgcactattgttgagaggttaaatggctctctcccttctgtatctta 4927

|||||

Db 349 ATTTCTTCTGATGCTACTATTGTTGGAGAGTTAAATGGTCTCTTCCCTTTGTGTAATCTTA 290

|||||

QY 4928 cctagtgttactcctgggcaccccttaattcctcagagggtgctaaattgtctgccattaca 4987

|||||

Db 289 CCTAGTGTCTTACTCTGGGCACCCCTTAATCTTTCAGAGGTGCTAAATTTGCTGCCATTACA 230

|||||

QY 4988 ccagaaggatgcctctgataggagacaaccatgcaaaattgtgaaatagtccttgaagttc 5047

|||||

Db 229 CCAGAAGGATGCTCTGTATAGGAGGACACCATGCAAAATGTGAAATAGTCTCTGAAGTTC 170

|||||

QY 5048 ttggattactttacacctcagttattgttccagaagaattttctggcctttcattggcaa 5107

|||||

Db 169 TTGGATTACTTTACACCTCAGTATTGATTGTGCCAGAGATTTTCTGGCCTTCATGGCAA 110

|||||

QY 5108 tgaataatttaagaagaagaatttaagatttttaatttaataagagtggttatataaata 5167

|||||

Db 109 TGAAATTTTAAAGAGAGAGATTTAAAGTATTTTAAATTTTAAAGAGTGTTGTTATAAATA 50

|||||

QY 5168 atgtactgaattctttatcccatctttatcatcttccagttttttattaa 5216

|||||

Db 49 ATGTACTGAATCTTTATCCCATTTTATCATCTCTTCAGTCTTTTATTAA 1

|||||

RESULT 16

AA1016896/c

LOCUS

DEFINITION

Ou31c08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone

EST

17-MAR-1999


```
Db 14 AGAGTACTACATACAGATTGGAAATCTCTAAAGACCGGTACTACCACTCGTTCTTTCA 73
QY 3361 ctcaacctgacaagaatttcagtaaacacacaaattatctgcctccttgaaatgaatcat 3420
Db 74 CTCAACTGTGACAGAAATTCAGTAACACCAAAATATCTGCCCTCTGAAATGAATCAT 133
QY 3421 ttttctctctggaaacacacaaacacaaatgtcttagagctgtttaaacaagccattt 3480
Db 134 TTTTCACTCTCTGGAAACCTTAAACACCAATGTCTTAGAGCTGTTAACAGCCACTTT 193
QY 3481 catcagcaggcaagaatctcagaccacaaatcatcacgaatggaaactgtgaagcaatgcaa 3540
Db 194 CATCAGCAGGCAAGCAATCTCAGACCAAAATCATCACGAATGGAACCTGTAAAGCAATGCAA 253
QY 3541 gcaacagctcaaatccaaagctctcctggaagaataaaggaggagcttgatagttctgaaa 3600
Db 254 GCACAGCTCAAAATCCAAAGCTCTCCTGGAAGATTAAGGGGAGGCTTCATAGTTCTGAAA 313
QY 3601 tggatcagtgaaatgaagattacacaaatgtcttcacctttcccggggaaaaaaagtg 3660
Db 314 TGGATCAGTGAAATGAAGATTACACATGTCTTCACCTTTCGCCGGGAAAAAAGTG 373
QY 3661 acaagagacgactctgatcttgaagttgctgaattggagaacctagagggcaggaaaa 3720
Db 374 ACAAGAGAGAGGACTCTGATCTTGTAGGTCTGAATTTGGAGAGCCTAGAGGCAGGAAAA 433
QY 3721 aaacgc 3726
Db 434 AAACGC 439

RESULT 13
LOCUS AI680124/c 423 bp mRNA EST 15-DEC-1999
DEFINITION tw6501.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2264521 3',
mRNA sequence.
ACCESSION AI680124
VERSION AI680124.1 GI:4890306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 423)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Clone Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/hbrp/image/image.html
Insert length: 1158 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 403.
FEATURES
source 1..423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2264521"
/clone_lib="NCI_CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.45 kb. Life Technologies catalog #:
```

```
BASE COUNT 11541-018"
ORIGIN 50 a 100 c 90 g 183 t

Query Match 8.0%; Score 423; DB 23; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.5e-174;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3803 tcagcgaagtctgaaaagagccatcagcgttcagaatctgataacagcagtgccctga 3862
Db 423 TCAGCGAAGTCGGAAGAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGCCTGA 364
QY 3863 ggaataagagctcaaaagaatattagaaaaatgaagatgaacagaatactccgcaaa 3922
Db 363 GGAAGAGAGGCTCAAGAAAGATATATTAGAAAATGAAGATGAACAGAATAGTCGCCCAA 304
QY 3923 aaagggTaaagagcgcagccaccacaaacctcttggTggaggtacacacaaagaagagcc 3982
Db 303 AAAGGTTAAAGAGCGCGACCCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGGCC 244
QY 3983 acaatgaactctctaaaaaaggaagcaaaaaaaatctggacctccagcaccagagga 4042
Db 243 AACAAATGAAACTTCTAAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGAGGA 184
QY 4043 ggaggaagaagaagaacaaactgaaatacgaagaacagaagtcacaaagcaacagca 4102
Db 183 GGAGGAAGAAGAAGAACAAGTGGAAATACCGGAACAGAAGTCCAAAACCAACAGCA 124
QY 4103 ccgagtgcaaggagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4162
Db 123 CCGAGTCTCAAGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 64
QY 4163 cacacagtccacaccacagaaaagagcagcagcagcagcagcagcagcagcagcagcagc 4222
Db 63 CACACAGTCCACACCACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4
QY 4223 acc 4225
Db 3 ACC 1

RESULT 14
LOCUS AW770571
DEFINITION h186f11.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3008877 3',
similar to TR:O94237 O94237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA
sequence.
ACCESSION AW770571
VERSION AW770571.1 GI:7702614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 578)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Clone Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Possible reversed clone: similarity on wrong strand
Seq primer: -400P from Gibco
High quality sequence stop: 465.
```



```

QY 2729 gctggcacaagaaccctgtttaccatgaatcatcacattagaacaatatatcagctatgtgc 2788
|||||
Db 417 GCTGGCACAAGAACCCTGTGTACCATGAATCATCACATTAGAACAAATATCAGCTATGTGC 476
|||||
QY 2789 attagctatcaacgatgaatgctcatcaagttaagacaagtttggcccaagaacttcacaa 2848
|||||
Db 477 ATTAGTATCAACGATGAATGCTATCAAGTAAACAAGTGTGTGGCCAGAAACTTCACAA 536
|||||
QY 2849 aggccttcccgctttacg 2867
|||||
Db 537 AGGCCTTTCGCGTTTACGG 555
|||||

RESULT 11
AW976150 738 bp mRNA EST 02-JUN-2000
LOCUS EST388259 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION AW976150
VERSION AW976150.1 GI:8167375
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 738)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Seq primer: Forward.
Location/Qualifiers
1..738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGN"
/notes="Vector: pBluescriptSKm"
BASE COUNT 214 a 126 c 129 g 269 t
ORIGIN

Query Match 8.2%; Score 431; DB 122; Length 738;
Best Local Similarity 99.6%; Pred. No. 4.7e-178;
Matches 531; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4674 atgtttctgatttcgaagctgtgtatgctttttatctgcggctttaaactgaacagtac 4733
|||||
Db 1 ATGTTTCTGATTCTGAAGTCTGTGTATAGCTTTATCTCGGCTTTAACTGACAGTAC 60
|||||
QY 4734 ccgactgtttattggatctattgattgaaagaattttagtagatagatcttaagcagt 4793
|||||
Db 61 CCGACTGTTTATTGGATCATTTGATTGTTGAAGAATTGTTAGGATAGATCTTAACAGT 120
|||||
QY 4794 aatctgcagtggtttgatttatttctgcaatttctactgcaaaaaaatttgtttt 4853
|||||
Db 121 AATCTGTCAGTGTGTTGTTATTTGTTATTTCTGCAATTTTACTGTGAAAAAATTTGTTTT 180
|||||
QY 4854 caacaattggtatctttctgagtcactattttttagagagtttaaatggtctcttcc 4913
|||||
Db 181 CAACAATTGGTGTCATTTCTTGTATGTCACATATTTGTTGGAGAGTTAAATGGTCTCTCC 240
|||||
QY 4914 ctttggatcttaccatagttttactctcctgggacccttaattcttcagaggtgctaaat 4973
|||||
Db 241 CTTTGTGTATCTTACCTAGTGTGTTACTCTCTGGCACCCCTTAATCTTCAGAGGTGCTAAAT 300
|||||

```

```

QY 4974 tgctgccattacaccagaaggatgctctgtgataggaggacaaccatgcaaatgtgaaa 5033
|||||
Db 301 TGCTGCCATTACACCACAAGGATGCTCTGTATAGGAGGACAACCATGCAAAATTTGTGAA 360
|||||
QY 5034 tagtctgaagttcttgatttaacttacacctcagtatgtattgtcccaagaattttctg 5093
|||||
Db 361 TAGTCTGTGAAGTTCTTTGGATTACTTTACACCTCAGTATTGATTGTCCCAAGATTTTCTG 420
|||||
QY 5094 gcctttcatggcaatgaaaattttaagaagaagatttaaaagtatttttaatttaagaag 5153
|||||
Db 421 GCCTTTTCATGGCAATGAAATTTTAAGAGAGAAGATTTAAAGTATTTTAAATTTAAAGAG 480
|||||
QY 5154 tegtgtataaaataatgtactgaattctttatccccattttatcatcctttcag 5206
|||||
Db 481 TGTGTTATAAAATAATGTACTGAATCTTTATCCCATTTTATCATCTCTTTTCAG 533
|||||

RESULT 12
AA282190 470 bp mRNA EST 13-AUG-1997
LOCUS zs89b08.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:704631 5',
DEFINITION mRNA sequence.
ACCESSION AA282190
VERSION AA282190.1 GI:1925314
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncilogap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 429.
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="GDB:585435"
/db_xref="taxon:9606"
/clone_lib="IMAGE:704631"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
cDNA was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTCTTTTCTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 169 a 101 c 95 g 105 t
ORIGIN

Query Match 8.1%; Score 426; DB 4; Length 470;
Best Local Similarity 100.0%; Pred. No. 7.4e-176;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3301 agagtactacatcagtttggaaatctcttaagaccgggtactaccagctcgtttcttca 3360
|||||

```


Db 122 GTATTGTTTAAAGAGTGTGTTATATAAATAATGTAATCTTCTTATCCCAATTTTA 63

Qy 5195 tcatacttccagtttttttataatc 5218
 |||||||
 Db 62 TCATCTCTTCAGTCTTTTATTATC 39

RESULT 10
 BE873840
 LOCUS BE873840 848 bp mRNA EST 20-OCT-2000
 DEFINITION 601483992F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3886511 5',
 mRNA sequence.
 ACCESSION BE873840
 VERSION BE873840.1 GI:10322616
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 848)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: DCTD/BTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LAM9663 row: g column: 24
 High quality sequence stop: 654.
 Location/Qualifiers
 1..848
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3886511"
 /clone_lib="NIH_MGC_69"
 /tissue_type="large cell carcinoma, undifferentiated"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."

BASE COUNT 261 a 163 c 191 g 232 t 1 others
 ORIGIN

Query Match 8.3%; Score 439; DB 141; Length 848;
 Best Local Similarity 100.0%; Pred. No. 1.4e-191;
 Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps

Qy 2429 ggtagctacttcattgtgaaagatcttctcatgaatgatcgctccagggaagaaagac 2488
 |||||||
 Db 117 GGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAATGATCGCTCCAGGGAAGAAC 176
 |||||||

Qy 2489 aactaaacttgggttccagatgaagaagatctctctgagacaattggtcaaatccaggc 2548
 |||||||
 Db 177 AACTAACTTTGGTTCAGATGAAGAGTATCTCTGAGACAATGGTCAAAATTCAGGC 236
 |||||||

Qy 2549 tattaaaaatgatggttcagatgggtcacttggaatgaaataataatcagtaaatcaggac 2608
 |||||||
 Db 237 TATTAAATGATGGTTCGATGGCTACTTTGGAATGAAATAATCATCACTAAATCAGGAAC 296
 |||||||

Qy 2609 ttctacctaaagatgtctaaacaaatttgcatagtgatggagactgacagaacaggg 2668
 |||||||
 Db 297 TTCTACCTTAAAGTTGCTAACACAATATTGATAGTATGAGAGACTTGACAGACAGG 356
 |||||||

Qy 2669 gaaaattagtaaacagatatgtcactgtctgagacttgcgtcgtggagtcgtattgtgaa 2728
 |||||||
 Db 357 GAAATTAAGTAACAGATATGTCAGCTGAGACTTGCTGCTGGAGTGCATTGTGTGAA 416
 |||||||

EST.
 human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 542)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 464.
 Location/Qualifiers
 1..542
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2813788"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: p7T7D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19w, testis NHT, and B-cell
 NCI-CGAP-GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 208 a 93 c 84 g 157 t

BASE COUNT 208 a 93 c 84 g 157 t
 ORIGIN

Query Match 8.6%; Score 453; DB 114; Length 542;
 Best Local Similarity 99.8%; Pred. No. 1.1e-187;
 Matches 503; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4715 ggctttaactgcagtcaccgactgtttattggactctattgattgaaagaattgtt 4774
 |||||||
 Db 542 GGCTTTAACTGACGATACCGACTGTTTATTCGACTATGATTGAAAGAAATTTGTT 483
 |||||||

Qy 4775 aggatagatttaagcagtaactgtcagtggtttgtattgtatttcttgcgaatttact 4834
 |||||||
 Db 482 AGGATAGATCTTAAGCAGTAATCTGTCAGTGTGTTGTTATTTGTTATCTCTGCAATTTTACT 423
 |||||||

Qy 4835 gtgaaaaaaattgttttcaacaattggtgcatttctcttgatgcatctattgttga 4894
 |||||||
 Db 422 GTGAAAAAAATTTGTTTTCAACAATTTGGTGTCAATTTCTTGATGTCATATTGTTGGA 363
 |||||||

Qy 4895 gagttaaagtgctctcccttctgtgtattctacactagtttactcctgggcacctta 4954
 |||||||
 Db 362 GAGTTAAATGTCCTCCCTTTGTGTATCTTACCTAGTGTGTTTACTCCTGGGCACCTTA 303
 |||||||

Qy 4955 atcttcagagtggtctaaattgtctgccattacaccagaagatgcctctgataggagac 5014
 |||||||
 Db 302 ATCTTCAGAGTGCTAAATGTCCTGCCATTACACCAGAAGGATGCCTTCGATAGGAGGAC 243
 |||||||

Qy 5015 aacatgcgaattgtgaaatagtcctgaagtcttcttggaacttctacacctcagttga 5074
 |||||||
 Db 242 AACATGCAAAATTTGTGAAATAGTCTCCTGAAGTTCTTGGATTACTTTACACCTCAGTATTGA 183
 |||||||

Qy 5075 ttgtccagaatttctggccttctatgcgaatgaaatttttaagaagaatttaa 5134
 |||||||
 Db 182 TTGTGCCAGAAATTTCTGCCCTTTCATGCGCAATGAAATTTTAAAGAAGAAATTTAA 123
 |||||||

Qy 5135 gtattttaatttaaaagtggtgtataaaataatgactgaattcttccatttca 5194
 |||||||

NCI CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 82 a 191 t 84 g 193 c

Query Match	8.7%	Score 460;	DB 23;	Length 480;
Best Local Similarity	100.0%;	Pred. No. 9e-191;		
Matches 460;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 3525	actgtaaagcaatgcaagcagcagcgtc	caaatccaagctctccttggaagaataaaagggaggg	3584	
DB 480	ACTGTAAGCAATGCAACGACGAGCTCA	AAATCCAAGCTCTCCTGGAAGAAATAAAGGGGAGG	421	
QY 3585	cttgatagttctgaaatggatcacagtg	aaaaatgaaagtatacaaatgtcttcacaccttg	3644	
DB 420	CTTGATAGTTCTGAAATGGATACAGT	GAAATGAAGATTACACAATGTCTTCCACCTTTG	361	
QY 3645	ccggggaaaaaaagtgcgaagagagacg	actctgtatcttgtaaagttctgaaatggagaag	3704	
DB 360	CCGGGGAAAAAAGTGACAGAGAGACG	ACTCTGTATCTTGTAAAGTCTGAATTCGAGAAG	301	
QY 3705	cctagagcgagggaaaaaaacgcgcgt	caacagacagagagagaaattaggtatgtagac	3764	
DB 300	CCTAGAGCGAGGAAAAAAGACCCGCT	CACAGACAGAGGAGAGAAATTTAGGTATGGATGAC	241	
QY 3765	ttagctaatgtgttacaggaacagaaac	cttaagcgagctcagcgaaatcggaagaagggc	3824	
DB 240	TTGACTTAAGTTGTTACAGGAAACAG	AAACCTTAAGGCAGTCAGCGAAGTCGGAAAGAGGC	181	
QY 3825	catacggtctcagaatctgatgaacagc	agctggccttgaggaagaagaggtctcaagaagaat	3884	
DB 180	CATACGCGTTTCAGAAATCTGATGAC	ACAGCAGCTGGCGCTCAGGAAAAAGAGGCTCA	121	
QY 3885	atattgaaaaatgaagatgaacagaat	tagtcgcgcacaaaaggggttaaaagagggccgacca	3944	
DB 120	ATATTAGAAATGGAAGATGAACAGAA	TATGTCCGCCAAAAAAGGGTTAAAGAGGCCGACCA	61	
QY 3945	ccaaaaccttgggtggaggtacaccca	aaaaagaagagccaa 3984		
DB 60	CCAAAACCTCTTGTTGGAGGTACAC	CAAAAAGAGAGGCCAA 21		

RESULT	8	
BF002153/c		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		

found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 463.
Location/Qualifiers
1. .540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:3316798"
/clone_lib="NCI_CGAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP.Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1057416-1061255, and 114584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo. "
209 a 93 c 85 g 153 t

Query Match	8.7%	Score 457;	DB 143;	Length 540;
Best Local Similarity	99.8%	Pred. No. 1.9e-189;		
Matches 507;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy 4711	ctcgcgtttaaactacagcagcaccgactcttatctgagctctattgattgaaagaatt	4770		
Db 540	CTGCGGCTTTAAACTGACAGTACCGGACTGTTATTGGATCTATGTGATTTGAAAGAATT	481		
Qy 4771	tgttaggatagactcttaagcagtaactctgcagtggttgattttgtatttctgtcaattt	4830		
Db 480	TGTTAGGATAGACTTTAAGCAGTAATCTGTCAGTGTGTGATTGTGATTCTGCAATTT	421		
Qy 4831	tactgtgaaaaaaatttgttttboaacaattggtgtgctatttctctgtatgctactattgt	4890		
Db 420	TACTGTGAAAAAAATTTGTTTTCACAAATTGGTGTGTCATTTTCTTGATGTGACATATTGT	361		
Qy 4891	tggagagttaaatggctctctcccttggctatctacctagtgtttacctcctggggcacc	4950		
Db 360	TGGAGAGTTAAATGGTCTCTCCCTTGTGTATCTTACCTAGTGTGTTTACCTGGGGCACC	301		
Qy 4951	cttaactctcagaggtgctcaaatgtctcgcattacaccagaagatgcctctgtatgga	5010		
Db 300	CTTAACTCTCAGAGGTGCTAAATGTCTGCCATTACACGAGGATGCCCTGATAGGA	241		
Qy 5011	ggacaaccatgcgaattgtgaaatagtcctggaagtcttggattactttacaactcagta	5070		
Db 240	GGACAACCATGCAAAATGTGAAATAGTCCCTGAAGTCTTGGATTACTCTTACCTCAGTA	181		
Qy 5071	ttgaattgtcccagaatttctcgcccttccatggccaatgaaaatttttaagaagaagatt	5130		
Db 180	TTGATTGTGCCAGAAATTTCTTGCCCTTTCATGGCAATGAAATTTTAAAGAAGAAGATT	121		
Qy 5131	taaaagtatttaattttcaagagtggtgtataaaaaaatgtaactgaaattctttatcccat	5190		
Db 120	TAAAGATTATTTTAAAGAGTGTGTTATAAATAAATGTAAGTACTGAATTTCTTATCCCAT	61		
Qy 5191	tttatacctcttcgaagtcttttataatc	5218		
Db 60	TTTATCATCTCTTCAGTTTTTATTAATC	33		

RESULT	9
AW303630/c	
LOCUS	
DEFINITION	542 bp mRNA EST
	xv2te03.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
	IMAGE:2813788 3', mRNA sequence.
ACCESSION	AW303630
VERSION	AW303630.1 GI:6713319

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 648)
AUTHORS
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL
Unpublished (2000)
COMMENT
The Institute for Genomic Research
Contact: John Quackenbush
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 317
Seq primer: Forward.

Location/Qualifiers
i..648
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE Resequences, MAGM"
/note="Vector: pBluescriptSKm"

BASE COUNT 257 a 112 c 94 g 185 t

ORIGIN

Query Match 10.2%; Score 536; DB 122; Length 648;
Best Local Similarity 99.8%; Pred. No. 4e-224; Indels 0; Gaps 0;
Matches 586; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4665 tttttttatgtttctgattgtgagtgctgtgtagcttttcttctggtctttaaagc 4724
Db 587 TTTTTTTTAAATGTTCTGATTTCTGAAGTCTGTGTAGCTTTATCTGCGGCTTTAAAC 528
QY 4725 tgacagaccgacttttattgagctctattgatttgaaagaatttggtaggagatc 4784
Db 527 TGACAGTACCCGACTCTTTATTGGATCTATTGATTTGAAAGAATTTGTTAGGATAGATC 468
QY 4785 ttaagcagtaactgtcagtggttattgttattttctgcaatttctgtaaaaaaa 4844
Db 467 TTAAGCAGTAATCTGTCAGTGTGTGTTGTTGATTTCTGCAATTTTACTGTGAAAAAA 408
QY 4845 attgttttcaacaattggtgtctattttcttgatgtcactatttggtagaggttaaatg 4904
Db 407 ATTTGTTTCAACAATGTTGTTCTGATTTCTTGATGTCACTATTGTTGGAGAGTTAAATG 348
QY 4905 gtctctccctttgtatcttaccctagtttacttccctgggacaccttaatttcagag 4964
Db 347 GTCTCTCCCTTTGNGTATCTTACCTAGTGTGTTTACCTGGGACCCCTTAATCTTCAGAG 288
QY 4965 gtgctaaattgtctgcattacaccagaaggatgctctgtagaggagacaaccatgcaa 5024
Db 287 GTGCTAAATGTCTGCCATTACACAGAGAGGATGCTCTGATAGGAGGACACCATGCAG 228
QY 5025 atttgaaatagtcctcgaagtcttggattactttacacctcagttattgattgcccag 5084
Db 227 ATTGTGAAATAGTCCCGAAGTCTTGGATTACTTTTACACCTCAGTATTGATTTGCCCCAG 168
QY 5085 aattttctgaccttccatggcaatgaaattttaagaagaagatttaagatttttaatt 5144
Db 167 AATTTTCTGGCCCTTCATGGCAATGAAATTTTAAAGAAGAAATTTAAAGTATTTTAAAT 108
QY 5145 ttttaagagtggtgtataaaataagtactgaatttttattccatttttccatttccttc 5204
Db 107 TTTAAAGAGTGTGTTTAAATAATCTACTGAATCTTTATCCATTTTATCATCTTTC 48
QY 5205 agtttttataactctactgtatcaataaaattctgtaatttgatga 5251
Db 47 AGTTTATTATTAATCTACTGTATCAATAAAAAATCTGTAATTTGAATGA 1

RESULT 5

AW629426/c

LOCUS

DEFINITION

hi57b12.x1 Soares NFL_T_GBC_S1 Homo sapiens cDNA clone

IMAGE:2976383 3', mRNA sequence.

ACCESSION

AW629426

VERSION

AW629426.1

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 568)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 444.

Location/Qualifiers

i..568

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NbHL19W, Testis NHT, and B-cell

NCI-CGAP GC81) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 220 a 97 c 87 g 164 t

ORIGIN

Query Match 9.7%; Score 511; DB 118; Length 568;

Best Local Similarity 99.8%; Pred. No. 3.8e-213;

Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4687 ctgaagtgcttgatagcttttctgctgctttaaactgacagtagctgctttatt 4746

Db 568 CTGAAGTGCTTGATAGCTTTTATCTGCGGCTTTAAACTGACAGTACCCGACTGTTATT 509

QY 4747 ggatctattgattgaaagaatttggtaggagatctttaaagcagtaattctgagtg 4806

Db 508 GGATCTATTGATTTGAAAAGAAATTTGTTAGGATAGATCTTTAAGCAGTAATCTGTCAGTGT 449

QY 4807 ttgtattgtatttctgcaattttactgtgaaaaaaattgttttcaacaattggtgt 4866

Db 448 TTGTATTGTATCTCTGCAATTTTACTGTGAAAAAAATTTGTTTCAACATTTGGTGT 389

QY 4867 cattttcttgatgctactatttggtagagatttaaatgctcttccctttgtagtctt 4926

Db 388 CATTTTCTTGATGTCATCTATTGTTGGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTT 329

QY 4927 acctagtgttactctggtgacaccttaattcttcagaggtgctaaattgctccattac 4986

Db 328 ACCTAGTGTTTACTCTCTGGCACCCCTTAATCTTTCAGAGGTGCTAAATTTGTGCGCATAC 269

QY 4987 accagaagatgctctctgtaggagagacaaccatgcaaatgtgaaatagctcgaagt 5046

Db 268 ACCAGAGGATGCCTCTGATAGGAGGACACCATGCAAAATTTGAAATAGTCTCTGAAGTT 209

ORIGIN

Query Match 11.3%; Score 596; DB 136; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.9e-250;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2752 atgaatcatcattagaaacattatcagctatgtcattagctatcaacgatgaatgct 2811
|||||
Db 1 ATGAAATCATCATATAGAACATATCATGCTATGTGCTATGCTATCAACGATGAATGCT 60
|||||

QY 2812 atcaagtaagacaagtgtttgcccagaaacttcacaaagccctttcccggttaccgcttc 2871
|||||
Db 61 ATCAAGTAAGACAAGTGTGTCGCCAGAACTTCACAAAGCGCTTTCGCCGTTTACGGCTTC 120
|||||

QY 2872 cacttgagtatatgcaaatctgtgcccctttgtgcaaaagatcctgttaaaagagagaag 2931
|||||
Db 121 CACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAG 180
|||||

QY 2932 ctcatgtaggcaagtgttggtgaaaaataataatgtaagcgggagtatctgaagcagc 2991
|||||
Db 181 CTATGCTAGGCAATGTTGGTGAAAAATATTAATGTAAAGCGGGAGTATCTGAAGCAGC 240
|||||

QY 2992 atgcagctgttagtgaataattattgtctctctaccagagtagttgtttccatatcaaa 3051
|||||
Db 241 ATGCAGCTGTAGTGAATAATTATTGTCTCTCTTACCAGAGTATGTTGTTCCATATACAA 300
|||||

QY 3052 ttcaacttttggcacatgacccagattatgtcaaaagtacaggatattgaacacattaaag 3111
|||||
Db 301 TTCACTTTTGGCACATGACCCAGATTATGTCAAAAGTACAGGATATTGAACAATTAAAG 360
|||||

QY 3112 atgttaagaagtcttctgtgttctgaaattatttaattgcttaaaatgaaataaaca 3171
|||||
Db 361 ATGTTAAAGAATGCTTTGTTGTTCTGGAATATTATTAATGGCTTAAATGAATAATCA 420
|||||

QY 3172 gtcaagcttttatcagaaagatgttagaaataatttaacaaacaaagagtgcccaaggac 3231
|||||
Db 421 GTCACGCTTTTATCAGAAAGATGTAGAAAAATATTAAACAAACAAAGATGCCCAAGGAC 480
|||||

QY 3232 cagatgtgcaaaaataaataagaaactgtacactgtgtgtgtgtgtggtgcaatgataca 3291
|||||
Db 481 CAGATGATGCAAAAATGAATGAAAACTGTACACTGTGTGTGTGTGTGTCATGAATATCA 540
|||||

QY 3292 tcatgtcaaaagagtactacatacagatttggaaatctctcaagaccggtactacca 3347
|||||
Db 541 TCATGTCAAGAGTACTACATACAGTTTGGAACTCTCTAAAGACCCCGGTACTACCA 596
|||||

RESULT 3
BG256731
LOCUS 734 bp mRNA EST 13-FEB-2001
DEFINITION 602370939F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4479153 5',
mRNA sequence.
ACCESSION BG256731
VERSION BG256731.1 GI:12766547
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 734)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-f@mail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM10310 row: m column: 10
High quality sequence stop: 732.

FEATURES
source Location/Qualifiers
1..734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4479153"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 278 a 154 c 161 g 141 t
ORIGIN

Query Match 10.9%; Score 576; DB 175; Length 734;
Best Local Similarity 100.0%; Pred. No. 1.1e-241;
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3318 ttggaatctcttaagaccggtgactaccagctggtttcttctcactcaacctgacaaagt 3377
|||||
Db 21 TTGGAATCTCTTAAGACCCGGTACTACCAGCTCGTTCTTCACTCAACCTGACAGAAT 80
|||||

QY 3378 ttcaagtaaccccaaaattatctgcctcctgaaatgaaatcaattttcactcctcggaaaa 3437
|||||
Db 81 TTCAGTAACACCAAAATTTATCTGCCCTCCTGAAATGAATCATTTTCACTCTCTGAAAA 140
|||||

QY 3438 cctaaaaacaaatgtcttagagctgttaaacagccactttcatcagcagcagaacaa 3497
|||||
Db 141 CCTAAAAACCAATGTTCTAGGAGCTGTTAAACAGGCCACTTTCATCAGCAGGCAAGCAA 200
|||||

QY 3498 tctcagaccataatcatcacgaatggaactgaaagcaatgcaagcagcagctcaaatcca 3557
|||||
Db 201 TCTCAGACCAATCATCACGAATGGAATCTGAAGCAATGCAACGACGAGCTCAATCCA 260
|||||

QY 3558 agctcctctggaagataaaagggaggcttgatgtcttctgaaatggatcagcagtgaaaaat 3617
|||||
Db 261 AGCTCTCTCTGGAAGATAAAGGGGAGGCTTGTATAGTTCTTGAATGGATCAGAGTGAAT 320
|||||

QY 3618 gaagattacaaatgtctcacccttgcgggggaaaaaagtacaaagagagacgactct 3677
|||||
Db 321 GAAGATTACAAATGCTCTTCCCTTTGCCGGGAAAAAAGTAGAACAGAGACGACTCT 380
|||||

QY 3678 gatctgttaaggtctgaattggagaagccttagagcaggagaaaaaacgcccgtcacagaa 3737
|||||
Db 381 GATCTTCTAAGGCTGAATTGGAGAAGCCTAGAGGCGAGGAAAAAACGCCCGTCCACAGAA 440
|||||

QY 3738 caggagagaaaaatagtgatggtatgactgaattggttacagggaacagaaacctaaa 3797
|||||
Db 441 CAGGAGGAGAAATTTAGGTATGATGATGACTTGAATGTTGGTACAGGAACAGAAACCTAAA 500
|||||

QY 3798 ggcagtcagcgaagtgcgaaaaagagcgcatacgtctcagaatctgatgaacagcagtg 3857
|||||
Db 501 GGCAGTCAGCAAGTCCGAAAAAGAGGCCATACGGCTTCAGAACTCTGATGAACAGCAGTGG 560
|||||

QY 3858 cctgagaaaaagaggtccaagaagatatattagaa 3893
|||||
Db 561 CCTGAGGAAAAAGAGGCTCAAGGAAGATATATTAGAA 596
|||||

RESULT 4
AW973437/c
LOCUS 648 bp mRNA EST 02-JUN-2000
DEFINITION EST385430 MAGE resequences, MAGM Homo sapiens cDNA, mRNA sequence.
ACCESSION AW973437
VERSION AW973437.1 GI:8164508
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 791)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES Location/Qualifiers
 1..791

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3086220"
 /clone_lib="NCI CGAP Sub8"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; NCI-CGAP-Sub8
 is a subcloned library derived from NCI-CGAP-Sub5. The
 NCI-CGAP-Sub8 library had 2.5 million recombinants. A
 single-stranded DNA preparation of NCI-CGAP-Sub5 was used
 as a tracer in a subtractive hybridization with a driver
 comprising a pool of clones from NCI-CGAP-Sub5 (IMAGE
 clone ids 2732833-2737415, 3068040-3069191; 25% of the
 driver population), a pool of clones from NCI-CGAP-Sub4
 (IMAGE clone ids 2723592-2729326; 25% of the driver
 population), NCI-CGAP-Sub6 (pool AIF-AJU, IMAGE ids
 2728969-2733190; 25% of the driver population), and
 NCI-CGAP-Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550
 ; 25% of the driver population). Subtraction was
 performed as previously described [Bonaldo, Lennon &
 Soares (1996): Normalization and Subtraction: Two
 Approaches To Facilitate Gene Discovery. Genome Research
 6: 791-806.
 TAG_LIB=NCI-CGAP_Lu5
 TAG_TISSUE=Lung
 TAG_SEQ=CAAC"

BASE COUNT 270 a 151 c 144 g 226 t
 ORIGIN

Query Match 12.0%; Score 631; DB 149; Length 791;
 Best Local Similarity 100.0%; Pred. No. 8.4e-266;
 Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2745 ttttaccatgaatcattacattagaaataatcaactatgctcattagctatcaacgat 2804
 DB 13 TGTACCAATGAATCATCATATAGAACCAATATACGCTATGTGCTATCATCAACGAT 72
 QY 2805 gaatgctatcaagtaagacaagtgtttgccagaaacttcacaaagccctttccqttta 2864
 DB 73 GAATGCTATCAAGTAGACAGAGTGTGCCAGAAATTCACAAAGCCCTTCCCGTTTA 132
 QY 2865 cggctccacttgagtataggaatctgtgccctttgtgcgaaagatcctgtaaaggag 2924
 DB 133 CGGCTTCCACTTGAATATAGCAATCTGTGCCCTTGTGTGCAAAAGATCCTGTAAAGGAG 192
 QY 2925 aaaaagctcattaggaatgtttgggaaaaataataatgtaagcgggagatctcg 2984
 DB 193 AGAAGAGCTCATCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGCGGGAGTATCTG 252
 QY 2985 aagcagatcagctgttagtgaataattattgtctctctaccagagatgttgttcca 3044
 DB 253 AGCAGCATGCAGCTTGTAGTGAATAATATTGTCTCTCTACCAGATGTGTGTCCA 312

QY 3045 tatacaattcaacttttggcacatgaccagattatgtcaaaatcacagattatgaacaa 3104
 DB 313 TATACAATTCACCTTTTGGCACATGACCAGATATATGTCAAAGTCAGAGATATGAACA 372
 QY 3105 cttaaagatgttaaagaatgtctttgtttgtttcttgaaatatttaagtgttaaaatgaa 3164
 DB 373 CTTAAAGATGTTAAAGAATGCTCTTGGTTGTCTGGAATATTAATGCTAAAAATGAA 432
 QY 3165 ataacagtcacgcctttatcagaaagatggttagaaatattatacaaaacaaagaatgcc 3224
 DB 433 AATAACAGTCACGCTTTTATCAGAAAGATGTTAGAAAATATTAAACAAACAAAGATGCC 492
 QY 3225 caaggaccagatgatcaaaaatgaatgaaactgtacactgtgtgtatgttgcacatg 3284
 DB 493 CAAGGACCAGATGATGCAAAAATGAATGAAAACCTGTACACTGTGTGTATGTTGCCATG 552
 QY 3285 aatatcatcatgcaaaagatcactacacagtttggatctctctaaagaccggtacta 3344
 DB 553 AATATCATCATGTCAAAGAGTACTACATAGTTCGGAATCTCTTAAGACCCGGTACTA 612
 QY 3345 ccagctcgtttcttctcactcaacctgacaaga 3375
 DB 613 CCAGCTCGTTCCTTCACTCAACCTGACAAGA 643

RESULT 2

BE504550

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE504550 596 bp mRNA EST 04-AUG-2000
 hz58a05 x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3212144.3
 similar to TR:094237 094237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA
 sequence.
 BE504550
 BE504550.1 GI:9706958
 EST.
 human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 596)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

Seq primer: -40UP from Gibco
 High quality sequence stop: 458.
 Location/Qualifiers
 1..596

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3212144"
 /clone_lib="NCI-CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Plasmid DNA from the normalized
 library NCI-CGAP_Lu5 was prepared, and ss circles were
 made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (cloneIDs
 1414920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 211 a 103 c 117 g 165 t

871	26	0.5	245	105	AL366935	AL366935 MCBALLIB04	944	26	0.5	285	132	BB356786	BB356786	
872	26	0.5	246	20	AI447785	AI447785 mq8ae01.x	945	26	0.5	286	27	AV258498	AV258498	
873	26	0.5	247	28	AV282689	AV282689 AV282689	946	26	0.5	287	118	AV633095	AV633095	
874	26	0.5	248	118	AW638728	AW638728 bl73f04.w	947	26	0.5	288	6	AA358545	AA358545	
875	26	0.5	249	188	T06477	T06477 EST04366.Fe	948	26	0.5	289	27	AV269715	AV269715	
876	26	0.5	250	119	AW641529	AW641529 cm08d09.w	949	26	0.5	290	118	AW633122	AW633122	
877	26	0.5	251	249	BF991574	BF991574 MR3-GN014	950	26	0.5	291	288	AQ093568	AQ093568	
878	26	0.5	252	249	A2846593	A2846593 2M0146C23	951	26	0.5	292	288	A2346917	A2346917	
879	26	0.5	253	4	AA237598	AA237598 mx11h04.r	952	26	0.5	293	114	AW305335	AW305335	
880	26	0.5	254	22	AI612112	AI612112 tt58b12.x	953	26	0.5	294	256	B33481	B33481	
881	26	0.5	255	111	AW088375	AW088375 x09e08.x	954	26	0.5	295	164	BE180440	BE180440	
882	26	0.5	256	139	BE721605	BE721605 189191.MA	955	26	0.5	296	103	AI925988	AI925988	
883	26	0.5	257	118	AW632874	AW632874 bl01c10.x	956	26	0.5	297	118	AW632841	AW632841	
884	26	0.5	258	244	AZ494100	AZ494100 1M0329H10	957	26	0.5	298	9	AA584275	AA584275	
885	26	0.5	259	19	AI335296	AI335296 ta47b03.x	958	26	0.5	299	118	AW635098	AW635098	
886	26	0.5	260	118	AW633139	AW633139 bl04c01.x	959	26	0.5	300	4	AA277233	AA277233	
887	26	0.5	261	188	R99438	R99438 YG79604.r1	960	26	0.5	301	118	AW633029	AW633029	
888	26	0.5	262	19	AI392442	AI392442 NCSC3C12T	961	26	0.5	302	118	AA260406	AA260406	
889	26	0.5	263	22	AI612027	AI612027 tt63d07.x	962	26	0.5	303	4	AA265180	AA265180	
890	26	0.5	264	22	AI621496	AI621496 486092B01	963	26	0.5	304	110	AV739720	AV739720	
891	26	0.5	265	27	AV264213	AV264213 AV264213	964	26	0.5	305	149	BF478508	BF478508	
892	26	0.5	266	113	AW263396	AW263396 qx8f02.x	965	26	0.5	306	14	AI003849	AI003849	
893	26	0.5	267	241	AZ311674	AZ311674 1M0027010	966	26	0.5	307	21	AI523898	AI523898	
894	26	0.5	268	3	AA212432	AA212432 mw76c09.r	967	26	0.5	308	109	AV044330	AV044330	
895	26	0.5	269	22	AI580595	AI580595 tb71q06.x	968	26	0.5	309	136	BE488154	BE488154	
896	26	0.5	270	102	AI862257	AI862257 tb90a02.x	969	26	0.5	310	118	AW634586	AW634586	
897	26	0.5	271	150	BF560275	BF560275 UI-R-E1-g	970	26	0.5	311	251	AZ883405	AZ883405	
898	26	0.5	272	165	BE234741	BE234741 141958.MA	971	26	0.5	312	118	AW633135	AW633135	
899	26	0.5	273	107	AU037846	AU037846 AU037846	972	26	0.5	313	170	BF853708	BF853708	
900	26	0.5	274	118	AW632889	AW632889 bl01e01.x	973	26	0.5	314	118	AW633135	AW633135	
901	26	0.5	275	118	AW632938	AW632938 bl02a03.x	974	26	0.5	315	300	173	EG087728	EG087728
902	26	0.5	276	19	AI344828	AI344828 ta99f10.x	975	26	0.5	316	170	BF853708	BF853708	
903	26	0.5	277	118	AW637298	AW637298 bl57a08.w	976	26	0.5	317	251	AZ883405	AZ883405	
904	26	0.5	278	171	BF930587	BF930587 MR2-NT013	977	26	0.5	318	118	AW633096	AW633096	
905	26	0.5	279	146	BF077836	BF077836 227994.MA	978	26	0.5	319	155	EG087728	EG087728	
906	26	0.5	280	144	BF077836	BF077836 227994.MA	979	26	0.5	320	170	BF874487	BF874487	
907	26	0.5	281	118	AW633097	AW633097 bl03g05.x	980	26	0.5	321	118	AW633135	AW633135	
908	26	0.5	282	121	AW834185	AW834185 MR2-TT001	981	26	0.5	322	118	AW633141	AW633141	
909	26	0.5	283	165	BE235913	BE235913 143540.MA	982	26	0.5	323	162	BB605008	BB605008	
910	26	0.5	284	110	AV738603	AV738603 AV738603	983	26	0.5	324	238	AZ108150	AZ108150	
911	26	0.5	285	113	AW264971	AW264971 qx58f10.x	984	26	0.5	325	18	AI267659	AI267659	
912	26	0.5	286	9	AA623752	AA623752 vG69b04.s	985	26	0.5	326	303	111	AW085910	AW085910
913	26	0.5	287	7	AA623752	AA623752 nc69f04.s	986	26	0.5	327	303	111	AW085910	AW085910
914	26	0.5	288	28	AV278690	AV278690 AV278690	987	26	0.5	328	303	121	AW868509	AW868509
915	26	0.5	289	27	AV277250	AV277250 AV277250	988	26	0.5	329	303	143	BF001154	BF001154
916	26	0.5	290	249	AV773912	AV773912 2M0001F13	989	26	0.5	330	160	BB520204	BB520204	
917	26	0.5	291	22	AI610952	AI610952 tt60f11.x	990	26	0.5	331	304	7	AA244626	AA244626
918	26	0.5	292	112	AW185154	AW185154 se87e10.y	991	26	0.5	332	304	7	AA244626	AA244626
919	26	0.5	293	240	AZ242469	AZ242469 RPT-23-7	992	26	0.5	333	304	27	AV258107	AV258107
920	26	0.5	294	118	AW633055	AW633055 bl03c08.x	993	26	0.5	334	151	BF592897	BF592897	
921	26	0.5	295	273	118	AW633146	AW633146 bl04c08.x	994	26	0.5	304	151	BF592897	BF592897
922	26	0.5	296	169	BF774517	BF774517 mx217.MA	995	26	0.5	305	162	BE014014	BE014014	
923	26	0.5	297	4	AA242620	AA242620 mx08d06.r	996	26	0.5	306	3	AA208851	AA208851	
924	26	0.5	298	111	AW104134	AW104134 xd5e03.x	997	26	0.5	307	122	AW971300	AW971300	
925	26	0.5	299	125	BB072629	BB072629 BB072629	998	26	0.5	308	1	AA034922	AA034922	
926	26	0.5	300	22	AI582986	AI582986 lr9a03.x	999	26	0.5	309	166	BE321246	BE321246	
927	26	0.5	301	164	BE148687	BE148687 MR0-HT024	1000	26	0.5					
928	26	0.5	302	27	AV265901	AV265901 AV265901								
929	26	0.5	303	118	AW633115	AW633115 bl03h12.x								
930	26	0.5	304	140	BE819309	BE819309 CMI-BN032								
931	26	0.5	305	279	AZ659664	AZ659664 1M0537M07								
932	26	0.5	306	18	AI287290	AI287290 qv60f02.x								
933	26	0.5	307	118	AW633036	AW633036 bl03b01.x								
934	26	0.5	308	281	118	AW633036	AW633036 bl03b01.x							
935	26	0.5	309	281	119	AW682376	AW682376 EST01185							
936	26	0.5	310	282	13	AA912454	AA912454 cm52b04.s							
937	26	0.5	311	282	19	AI349675	AI349675 tb58d06.x							
938	26	0.5	312	114	AW301766	AW301766 xr83d06.x								
939	26	0.5	313	282	118	AW633127	AW633127 bl04a12.x							
940	26	0.5	314	120	AW803530	AW803530 IL2-UM008								
941	26	0.5	315	230	AQ566683	AQ566683 HS_2104.B								
942	26	0.5	316	249	AZ781349	AZ781349 2M0019A09								
943	26	0.5	317	110	AV739714	AV739714 AV739714								

ALIGNMENTS

RESULT 1

BF509252

LOCUS

DEFINITION

UI-H-B14-aow-c-07-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens CDNA clone IMAGE:3086220 3', mRNA sequence.

ACCESSION

BF509252

VERSION

BF509252.1

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ALIGNMENTS

RESULT 1

BF509252 791 bp mRNA EST
UI-H-B14-aov-c-07-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3086220 3', mRNA sequence.
BF509252 GI:11592550
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

c 725	26	0.5	125	119	AW642236	AW642236 cml7d03.w	798	26	0.5	202	241	A231604	A231604	lmo059H07
c 726	26	0.5	131	18	AW6310949	AW6310949 ta76d08.x	799	26	0.5	203	164	BE150909	BE150909	RC4-HF027
c 727	26	0.5	133	118	AW639663	AW639663 bl85e10.w	c 800	26	0.5	204	22	AW588437	AW588437	fc02a06.x
c 728	26	0.5	136	19	AW6345273	AW6345273 tb68a10.x	c 801	26	0.5	204	118	AW633071	AW633071	bl103e02.x
c 729	26	0.5	139	118	AW633102	AW633102 bl03a10.x	c 802	26	0.5	204	118	AW633756	AW633756	bl11905.w
c 730	26	0.5	144	18	AW6310875	AW6310875 ta87a05.x	c 803	26	0.5	204	118	AW633766	AW633766	bl11h04.w
c 731	26	0.5	144	190	W39837	W39837 333 Mouse v	c 804	26	0.5	205	112	AW633066	AW633066	se87g03.y
c 732	26	0.5	145	165	BE235153	BE235153 142506 MA	c 805	26	0.5	205	112	AW633066	AW633066	bl03d08.x
c 733	26	0.5	146	3	AA209567	AA209567 mw75a07.r	c 806	26	0.5	206	103	AW633066	AW633066	bl03d08.x
c 734	26	0.5	146	137	BE1371638	BE1371638 tb72h09.x	c 807	26	0.5	206	118	AW633165	AW633165	bl04e03.x
c 735	26	0.5	146	122	AW946864	AW946864 RC2-EF002	c 808	26	0.5	206	170	BF890592	BF890592	CMI-MF010
c 736	26	0.5	147	118	AW632958	AW632958 bl02e01.x	c 809	26	0.5	208	19	AI340523	AI340523	tb30c07.x
c 737	26	0.5	148	3	AA207069	AA207069 zr87e06.s	c 810	26	0.5	209	20	AI342756	AI342756	th49b03.x
c 738	26	0.5	148	18	AI249411	AI249411 qx74g02.x	c 811	26	0.5	209	118	AW633025	AW633025	bl03a01.x
c 739	26	0.5	149	137	BE610701	BE610701 sg64g11.y	c 812	26	0.5	209	164	AW633025	AW633025	bl03a01.x
c 740	26	0.5	150	17	AI232458	AI232458 EST229146	c 813	26	0.5	209	118	BE151756	BE151756	QV0-HF030
c 741	26	0.5	150	19	AI345528	AI345528 tb69h03.x	c 814	26	0.5	210	22	AI613386	AI613386	bl08e12.x
c 742	26	0.5	151	118	AW633195	AW633195 bl04g10.x	c 815	26	0.5	211	18	AI270927	AI270927	qw52b09.x
c 743	26	0.5	154	118	AW633164	AW633164 bl04e02.x	c 816	26	0.5	211	171	BF909224	BF909224	PH3-UT005
c 744	26	0.5	154	137	BE548041	BE548041 601072072	c 817	26	0.5	212	118	AW633169	AW633169	bl04e08.x
c 745	26	0.5	155	17	AI223731	AI223731 qx32b03.x	c 818	26	0.5	212	132	BB371564	BB371564	BB371564
c 746	26	0.5	157	162	BE062038	BE062038 RC1-BT025	c 819	26	0.5	213	4	AA237192	AA237192	mx17h03.r
c 747	26	0.5	158	8	AA514673	AA514673 nf58e05.s	c 820	26	0.5	213	19	AI377146	AI377146	tb89e08.x
c 748	26	0.5	159	165	BE234108	BE234108 140717 MA	c 821	26	0.5	213	118	AW633186	AW633186	bl04g01.x
c 749	26	0.5	160	7	AA417829	AA417829 zv01a10.r	c 822	26	0.5	214	18	AI273880	AI273880	qu09d01.x
c 750	26	0.5	162	175	BG237164	BG237164 sab04d09.y	c 823	26	0.5	214	170	BF877366	BF877366	QV2-EF010
c 751	26	0.5	163	103	AI889818	AI889818 wm63f07.x	c 824	26	0.5	215	19	AI401159	AI401159	tg26f01.x
c 752	26	0.5	163	174	BG166238	BG166238 602345407	c 825	26	0.5	215	120	AW795810	AW795810	MR4-UM002
c 753	26	0.5	164	1	AA059132	AA059132 zf64c09.s	c 826	26	0.5	219	22	AI630306	AI630306	ad08c12.y
c 754	26	0.5	164	107	AU074500	AU074500 AU074500	c 827	26	0.5	219	110	AV739377	AV739377	AV739377
c 755	26	0.5	164	122	AW891165	AW891165 PM2-NF007	c 828	26	0.5	220	4	AA244598	AA244598	mx31n06.r
c 756	26	0.5	165	102	AI792457	AI792457 q168a09.y	c 829	26	0.5	221	104	AI990172	AI990172	ws38b01.x
c 757	26	0.5	165	111	AW501954	AW501954 UI-HF-BN0	c 830	26	0.5	221	114	AW301935	AW301935	xf85b01.x
c 758	26	0.5	169	67	DMAA40479	AA40479 LD12542.5	c 831	26	0.5	222	19	AI373773	AI373773	qx254f05.x
c 759	26	0.5	169	137	BE552916	BE552916 946087D01	c 832	26	0.5	223	18	AI310027	AI310027	tb49b10.x
c 760	26	0.5	171	102	AI799369	AI799369 tw33e10.x	c 833	26	0.5	223	121	AW856291	AW856291	RC1-CF028
c 761	26	0.5	171	130	BB277223	BB277223 BB277223	c 834	26	0.5	223	151	BF591791	BF591791	7150e05.x
c 762	26	0.5	172	7	AA430552	AA430552 zw22g06.s	c 835	26	0.5	223	155	BG610615	BG610615	727915 MA
c 763	26	0.5	172	118	AW633003	AW633003 bl02g03.x	c 836	26	0.5	224	17	AI242020	AI242020	qv51f10.x
c 764	26	0.5	175	18	AI3111636	AI3111636 ta43h10.x	c 837	26	0.5	224	22	AI580598	AI580598	tb71g09.x
c 765	26	0.5	175	103	AI906649	AI906649 RC-BT122-	c 838	26	0.5	224	23	AI675727	AI675727	wc40g04.x
c 766	26	0.5	176	169	BF814530	BF814530 MR2-CI018	c 839	26	0.5	225	12	AA809479	AA809479	ob85c02.s
c 767	26	0.5	176	110	AV740432	AV740432 AV740432	c 840	26	0.5	225	122	AW897965	AW897965	RC3-NN006
c 768	26	0.5	176	110	AV741164	AV741164 AV741164	c 841	26	0.5	227	22	AI580614	AI580614	tb72a05.x
c 769	26	0.5	176	110	AV741894	AV741894 AV741894	c 842	26	0.5	230	170	BF875604	BF875604	QV3-EF010
c 770	26	0.5	176	118	AW638291	AW638291 bl68c12.w	c 843	26	0.5	231	24	AI758760	AI758760	ty15h06.x
c 771	26	0.5	177	142	BE964812	BE964812 601658467	c 844	26	0.5	231	139	BE693049	BE693049	SAC503 MO
c 772	26	0.5	179	10	AA691308	AA691308 vs14e09.r	c 845	26	0.5	232	118	AW633106	AW633106	bl03h03.x
c 773	26	0.5	180	102	AI792533	AI792533 q177a12.y	c 846	26	0.5	233	114	AW301411	AW301411	x876f01.x
c 774	26	0.5	180	118	AW640434	AW640434 bl195805.w	c 847	26	0.5	233	120	AW748972	AW748972	RC1-BF031
c 775	26	0.5	181	22	AI612793	AI612793 tp44f02.x	c 848	26	0.5	233	172	BG004526	BG004526	ILO-GN028
c 776	26	0.5	183	18	AI284741	AI284741 qu24e01.x	c 849	26	0.5	233	240	AZ267979	AZ267979	RPCI-23-1
c 777	26	0.5	184	19	AI349605	AI349605 ta73g12.x	c 850	26	0.5	234	123	BB018787	BB018787	BB018787
c 778	26	0.5	185	18	AI289689	AI289689 gw42e10.x	c 851	26	0.5	235	3	AA208776	AA208776	mw73f02.r
c 779	26	0.5	186	3	AA212578	AA212578 mw77f09.r	c 852	26	0.5	235	18	AI306755	AI306755	qw44b05.x
c 780	26	0.5	187	118	AW633150	AW633150 bl04c12.x	c 853	26	0.5	235	19	AI369117	AI369117	qw43d05.x
c 781	26	0.5	187	122	AW892681	AW892681 CM3-NN000	c 854	26	0.5	235	20	AI3436486	AI3436486	th51f10.x
c 782	26	0.5	189	118	AW632950	AW632950 bl02b03.x	c 855	26	0.5	236	10	AI349974	AI349974	tb59a05.x
c 783	26	0.5	189	118	AW633207	AW633207 bl04h11.x	c 856	26	0.5	236	22	AI567673	AI567673	tr64b02.x
c 784	26	0.5	191	164	BE160102	BE160102 V01-HT041	c 857	26	0.5	236	123	BB016612	BB016612	BB016612
c 785	26	0.5	192	151	BF654195	BF654195 278266 MA	c 858	26	0.5	238	4	AA244555	AA244555	mx04h11.r
c 786	26	0.5	193	118	AW633185	AW633185 bl04f12.x	c 859	26	0.5	239	118	AW636849	AW636849	bl51d08.w
c 787	26	0.5	193	163	BE138942	BE138942 xr63c12.x	c 860	26	0.5	240	28	AV312773	AV312773	AV312773
c 788	26	0.5	194	12	AA822798	AA822798 vq91b11.r	c 861	26	0.5	240	103	AI880491	AI880491	ac78h07.x
c 789	26	0.5	195	118	AW633000	AW633000 bl02f12.x	c 862	26	0.5	240	107	AU073621	AU073621	AU073621
c 790	26	0.5	196	175	BG298023	BG298023 60236236	c 863	26	0.5	241	21	AI537464	AI537464	tp04b11.x
c 791	26	0.5	199	19	AI343135	AI343135 tb04f10.x	c 864	26	0.5	241	27	AV270039	AV270039	AV270039
c 792	26	0.5	200	115	AW426270	AW426270 60019 MAR	c 865	26	0.5	242	18	AI275677	AI275677	qv66g12.x
c 793	26	0.5	201	116	AW484318	AW484318 60018 MAR	c 866	26	0.5	243	118	AW633159	AW633159	bl04d09.x
c 794	26	0.5	201	118	AW633208	AW633208 bl04h12.x	c 867	26	0.5	243	172	BF989949	BF989949	MR2-GN012
c 795	26	0.5	201	122	AW905296	AW905296 QV2-NN107	c 868	26	0.5	244	118	AW632844	AW632844	bl01a04.x
c 796	26	0.5	202	18	AI269268	AI269268 q125e04.x	c 869	26	0.5	245	19	AI334840	AI334840	qw43e01.x
c 797	26	0.5	202	111	AW071272	AW071272 xa63c06.x	c 870	26	0.5	245	21	AI521049	AI521049	to70c04.x

579	27	0.5	514	20	AI436455	ti09c11.x	C 652	27	0.5	669	247	AZ648112	IM0517A05	
580	27	0.5	517	237	AZ072300	RPCI-23-3	C 653	27	0.5	670	248	AZ738195	RPCI-24-1	
581	27	0.5	520	242	AZ339006	IM0070L13	C 654	27	0.5	671	103	AI893653	mn18b05.x	
582	27	0.5	521	242	AZ368976	IM0119M17	C 655	27	0.5	672	241	AZ324179	IM0046F03	
583	27	0.5	525	14	AI011151	EST205602	C 656	27	0.5	673	233	AQ742061	HS_5570_B	
584	27	0.5	526	18	AI287470	q81e09.x	C 657	27	0.5	674	236	AQ088648	RPCI-23-3	
585	27	0.5	528	110	AV754096	AV754096	C 658	27	0.5	675	728	BF678033	602085072	
586	27	0.5	532	11	AA744512	ny65c07.s	C 659	27	0.5	676	168	BF681835	602117674	
587	27	0.5	533	110	AV755215	AV755215	C 660	27	0.5	677	233	AQ749077	HS_5538_A	
588	27	0.5	533	233	AQ781023	HS_3169_B	C 661	27	0.5	678	22	AI567787	ti62f01.x	
589	27	0.5	535	221	CNS04FGJ	Tetraodon	C 662	27	0.5	679	32	AV710961	AV710961	
590	27	0.5	535	242	AZ353018	AZ353018	C 663	27	0.5	680	107	AI213336	AU123336	
591	27	0.5	536	16	AI159622	uf02f05.x	C 664	27	0.5	681	146	BF245407	BF245407	
592	27	0.5	538	223	AQ043211	CIT-HSP-2	C 665	27	0.5	682	169	BF792701	602253895	
593	27	0.5	543	117	AM538870	C0112H04-	C 666	27	0.5	683	201	AI624300	ts28108.x	
594	27	0.5	546	167	BE430276	BE430276	C 667	27	0.5	684	22	AW332494	S9B AGS-	
595	27	0.5	547	23	AI651924	w51b10.x	C 668	27	0.5	685	114	AW332494	S9B AGS-	
596	27	0.5	547	246	AZ653047	RPCI-23-2	C 669	27	0.5	686	146	BF244960	601864371	
597	27	0.5	548	111	AM070218	xa09d02.x	C 670	27	0.5	687	219	AI053991	Drosophila	
598	27	0.5	548	156	C99229	C99229	C 671	27	0.5	688	123	AW983107	HVSMG000	
599	27	0.5	548	164	BE186233	BE186233	C 672	27	0.5	689	174	BG167235	602345386	
600	27	0.5	549	24	AI742699	wg43b03.x	C 673	27	0.5	690	839	BF107623	601823981	
601	27	0.5	549	121	AM840682	AM840682	C 674	27	0.5	691	144	BF107623	601823981	
602	27	0.5	550	21	AI503368	AI503368	C 675	27	0.5	692	135	BF678635	602085214	
603	27	0.5	553	118	AM634145	AM634145	C 676	27	0.5	693	872	AQ746568	HS_2278_A	
604	27	0.5	555	22	AI567549	tr89g08.x	C 677	27	0.5	694	222	CNS05F0E	Tetraodon	
605	27	0.5	555	172	BF696957	BF696957	C 678	27	0.5	695	80	BF271208	GA_Eb001	
606	27	0.5	556	251	AZ913961	RPCI-24-1	C 679	27	0.5	696	154	BG501720	602548777	
607	27	0.5	558	256	BE5614	BE5614	C 680	27	0.5	697	233	AQ742145	HS_5383_B	
608	27	0.5	562	112	AM184368	AM184368	C 681	27	0.5	698	233	AQ746591	HS_2278_A	
609	27	0.5	567	9	AA599483	AA599483	C 682	27	0.5	699	966	222	CNS05FAY	Tetraodon
610	27	0.5	568	136	BE473339	sp51a10.y	C 683	27	0.5	699	143	BF038819	601462003	
611	27	0.5	570	32	AV717482	AV717482	C 684	26	0.5	700	4	AA254893	AA254893	
612	27	0.5	570	241	AZ311236	AZ311236	C 685	26	0.5	701	244	AZ453639	IM0255008	
613	27	0.5	575	118	AM637920	AM637920	C 686	26	0.5	702	118	AM632856	b101b04.x	
614	27	0.5	579	114	AM331093	AM331093	C 687	26	0.5	703	53	AI254432	qV34005.x	
615	27	0.5	580	224	AQ078766	CIT-HSP-2	C 688	26	0.5	704	118	AM632959	AI254432	
616	27	0.5	583	241	AZ281083	RPCI-23-1	C 689	26	0.5	705	62	AI252328	qV27b08.x	
617	27	0.5	585	248	AZ736910	RPCI-24-7	C 690	26	0.5	706	118	AM633014	AI252328	
618	27	0.5	585	250	AZ843882	AZ843882	C 691	26	0.5	707	22	AI612006	tl63b03.x	
619	27	0.5	592	122	AQ149378	AQ149378	C 692	26	0.5	708	118	AM633089	b103f09.x	
620	27	0.5	593	112	AM192764	AM192764	C 693	26	0.5	709	118	AM632877	b101d01.x	
621	27	0.5	593	229	AQ488486	RPCI-11-2	C 694	26	0.5	710	18	AI254562	qV49a11.x	
622	27	0.5	594	235	AQ899648	HS_2046_B	C 695	26	0.5	711	18	AI250613	AI250613	
623	27	0.5	597	244	AZ468128	AM281A05	C 696	26	0.5	712	18	AI250615	qV33h11.x	
624	27	0.5	602	144	BF131953	BF131953	C 697	26	0.5	713	18	AI250615	qV33h11.x	
625	27	0.5	605	241	AZ333993	AZ333993	C 698	26	0.5	714	18	AI250615	qV33h11.x	
626	27	0.5	605	248	AZ741812	AZ741812	C 699	26	0.5	715	18	AI254572	qV49b11.x	
627	27	0.5	606	242	AZ378690	AM1331L04	C 700	26	0.5	716	18	AI254572	qV49b11.x	
628	27	0.5	606	243	AZ395910	AM160P15	C 701	26	0.5	717	18	AI254562	qV49a11.x	
629	27	0.5	607	241	AZ313543	AM0029P12	C 702	26	0.5	718	18	AI254562	qV49a11.x	
630	27	0.5	607	249	AZ793664	AZ793664	C 703	26	0.5	719	18	AI254562	qV49a11.x	
631	27	0.5	612	238	AZ121188	AZ121188	C 704	26	0.5	720	18	AI254562	qV49a11.x	
632	27	0.5	617	105	AL120173	AL120173	C 705	26	0.5	721	18	AI254562	qV49a11.x	
633	27	0.5	618	251	AZ874453	AZ874453	C 706	26	0.5	722	18	AI254562	qV49a11.x	
634	27	0.5	627	230	AQ552075	RPCI-11-4	C 707	26	0.5	723	18	AI254562	qV49a11.x	
635	27	0.5	628	244	AZ455001	AM0257M10	C 708	26	0.5	724	18	AI254562	qV49a11.x	
636	27	0.5	631	119	AW272071	GA_Ea001	C 709	26	0.5	725	18	AI254562	qV49a11.x	
637	27	0.5	631	226	AQ284603	AQ284603	C 710	26	0.5	726	18	AI254562	qV49a11.x	
638	27	0.5	632	242	AZ376810	AM1030P18	C 711	26	0.5	727	18	AI254562	qV49a11.x	
639	27	0.5	633	112	AM180978	AM180978	C 712	26	0.5	728	18	AI254562	qV49a11.x	
640	27	0.5	637	238	AZ107481	RPCI-23-3	C 713	26	0.5	729	18	AI254562	qV49a11.x	
641	27	0.5	639	20	AI445562	tj30g02.x	C 714	26	0.5	730	18	AI254562	qV49a11.x	
642	27	0.5	641	144	BF131261	BF131261	C 715	26	0.5	731	18	AI254562	qV49a11.x	
643	27	0.5	643	242	AZ376123	AM129P18	C 716	26	0.5	732	18	AI254562	qV49a11.x	
644	27	0.5	644	241	AZ336755	AM0067J08	C 717	26	0.5	733	18	AI254562	qV49a11.x	
645	27	0.5	648	249	AZ795850	AM0051J17	C 718	26	0.5	734	18	AI254562	qV49a11.x	
646	27	0.5	649	103	AI1889305	wm97c10.x	C 719	26	0.5	735	18	AI254562	qV49a11.x	
647	27	0.5	649	168	BF675817	BF675817	C 720	26	0.5	736	18	AI254562	qV49a11.x	
648	27	0.5	651	142	BE964060	BE964060	C 721	26	0.5	737	18	AI254562	qV49a11.x	
649	27	0.5	653	175	BE283218	BE283218	C 722	26	0.5	738	18	AI254562	qV49a11.x	
650	27	0.5	667	220	CNS027BZ	CNS027BZ	C 723	26	0.5	739	18	AI254562	qV49a11.x	
651	27	0.5	668	243	AZ422289	IM0200I11	C 724	26	0.5	740	18	AI254562	qV49a11.x	

433	10	AA701131	AA701131 z956d10.s	506	27	0.5	415	104	AJ284206	AJ284206 4A3B-AAU-
434	249	AE784921	AE784921 2M0028K14	507	27	0.5	419	19	A1333107	A1333107 qq26h08.x
435	319	BE702016	BE702016 OVA-BT040	508	27	0.5	420	136	BE502535	BE502535 hy16d04.x
436	163	BB093363	BB093363 BB093363	509	27	0.5	420	187	R42066	R42066 vF98b07.s1
437	321	AA794731	AA794731 RC6-UM001	510	27	0.5	421	2	AA088951	AA088951 mm51h08.r
438	326	AA093063	AA093063 RPCI-23-4	511	27	0.5	423	249	AZ778114	AZ778114 2M0013E12
439	327	AA1451720	AA1451720 mx26c09.x	512	27	0.5	425	17	A1236839	A1236839 EST33401
440	331	AA209171	AA209171 AV209171	513	27	0.5	425	23	A1656555	A1656555 ts1f05.x
441	331	AA133537	AA133537 DKF2P762E	514	27	0.5	425	168	BF690990	BF690990 602247022
442	332	AA1564682	AA1564682 tq78q08.x	515	27	0.5	426	141	BE889073	BE889073 601513512
443	332	BB159288	BB159288 BB159288	516	27	0.5	428	23	A1651930	A1651930 wb51c04.x
444	332	BB114478	BB114478 BB114478	517	27	0.5	430	15	A1051289	A1051289 ox24b02.x
445	334	BE032563	BE032563 131966.MA	518	27	0.5	430	23	A1650430	A1650430 wa90f01.x
446	334	BB233438	BB233438 dda28b11.	519	27	0.5	431	142	BE968824	BE968824 601649848
447	338	AA1580655	AA1580655 to72a09.x	520	27	0.5	433	104	AL022903	AL022903 t8130b51
448	339	BE416321	BE416321 MUG007.E0	521	27	0.5	438	20	A1451807	A1451807 mx10e11.y
449	340	BE232022	BE232022 136928.MA	522	27	0.5	438	104	A1954690	A1954690 wq32a07.x
450	340	BF986207	BF986207 RC3-GN004	523	27	0.5	438	234	AQ820872	AQ820872 HS_5263_B
451	342	AA055850	AA055850 CIR-HSP-2	524	27	0.5	439	118	AW632847	AW632847 b101a07.x
452	343	AA634283	AA634283 b118906.w	525	27	0.5	441	22	A1582556	A1582556 ts02e11.x
453	345	AA2468166	AA2468166 lM0281I04	526	27	0.5	442	150	BF551292	BF551292 UI-R-CO-h
454	346	AA219225	AA219225 zq16g12.r	527	27	0.5	444	17	A1186252	A1186252 qd34d02.x
455	346	AA707038	AA707038 sk30g08.y	528	27	0.5	444	136	BE519343	BE519343 945023C08
456	346	BB274694	BB274694 BB274694	529	27	0.5	445	149	BF469402	BF469402 UI-M-BH3-
457	348	BE665360	BE665360 154308.MA	530	27	0.5	446	118	AW632941	AW632941 b102a06.x
458	350	AA655964	AA655964 vs45a02.r	531	27	0.5	448	247	AZ639153	AZ639153 lM0499N21
459	350	AA1462298	AA1462298 ub62a10.x	532	27	0.5	451	16	A1138555	A1138555 qd80h08.x
460	350	BE120293	BE120293 UI-R-CA0-	533	27	0.5	451	22	A1586319	A1586319 vs45a02.x
461	352	AA280412	AA280412 zt05d02.s	534	27	0.5	451	103	A1912631	A1912631 we11h06.x
462	353	AA2116802	AA2116802 RPCI-23-4	535	27	0.5	451	249	AZ753044	AZ753044 RPCI-24-8
463	355	AA204375	AA204375 RPCI-23-3	536	27	0.5	453	9	AA618552	AA618552 np38d01.s
464	357	AA678291	AA678291 AV678291	537	27	0.5	455	13	AA938656	AA938656 oc08g01.s
465	362	AA1556038	AA1556038 UI-R-C2P-	538	27	0.5	455	145	BF193384	BF193384 249531.MA
466	363	AA209596	AA209596 mw75g08.r	539	27	0.5	456	137	BE575511	BE575511 dc57d01.x
467	364	AA279444	AA279444 z584g02.s	540	27	0.5	461	240	AZ246274	AZ246274 RPCI-23-3
468	366	AA2267412	AA2267412 RPCI-23-4	541	27	0.5	462	19	A1348771	A1348771 ta85a11.x
469	367	AA145699	AA145699 ga32h08.y	542	27	0.5	462	118	AW633057	AW633057 b103c10.x
470	370	AA710174	AA710174 el909ne.f	543	27	0.5	463	156	C22842	C22842 C22642.Rice
471	371	BF059743	BF059743 7k65904.x	544	27	0.5	463	187	N94183	N94183 za26d12.r1
472	373	AA589566	AA589566 xo12b08.x	545	27	0.5	465	23	A1650622	A1650622 wb01g08.x
473	376	AA721019	AA721019 nx89f10.s	546	27	0.5	468	250	AZ832378	AZ832378 2M0112B15
474	376	AA771567	AA771567 vm46b11.r	547	27	0.5	469	118	AW632912	AW632912 b101f12.x
475	376	BF131961	BF131961 601820940	548	27	0.5	470	32	AA715898	AA715898 AV715898
476	381	AA283388	AA283388 RPCI111-89	549	27	0.5	475	146	BF286258	BF286258 EST450849
477	383	AA1059875	AA1059875 UI-R-C1-1	550	27	0.5	476	237	AZ068159	AZ068159 RPCI-23-4
478	385	AA642708	AA642708 cm21f09.w	551	27	0.5	477	20	A1477812	A1477812 fb55g04.y
479	386	AA2753512	AA2753512 RPCI-24-1	552	27	0.5	478	110	AV754037	AV754037 AV754037
480	388	AA209579	AA209579 mw75g07.r	553	27	0.5	478	223	AQ040554	AQ040554 CIR-HSP-2
481	388	AA085780	AA085780 xc55g03.x	554	27	0.5	479	237	AZ051332	AZ051332 s1c00039
482	392	AA723324	AA723324 AV723324	555	27	0.5	481	143	BF059431	BF059431 7K359e07.x
483	390	AA678205	AA678205 z115c11.s	556	27	0.5	481	245	AZ517260	AZ517260 RPCI-11-8
484	391	AA766766	AA766766 oa35c10.s	557	27	0.5	483	189	W03733	W03733 za74e07.r1
485	395	AA213318	AA213318 mw83e06.r	558	27	0.5	484	16	A1126716	A1126716 qb94d01.x
486	395	AA292006	AA292006 z128b11.r	559	27	0.5	490	102	A1806046	A1806046 te48b01.x
487	397	AA601154	AA601154 no54f11.s	560	27	0.5	490	104	A1988136	A1988136 wu13e10.x
488	397	AA1267656	AA1267656 aq92g10.x	561	27	0.5	491	256	B23561	B23561 F15G15TF IG
489	401	AA1604666	AA1604666 vm46b11.y	562	27	0.5	492	116	AA435086	AA435086 UI-R-BJ0P
490	402	AA454486	AA454486 zw29e02.r	563	27	0.5	496	22	A1631896	A1631896 wa37d10.x
491	402	AA1467453	AA1467453 vd53g10.x	564	27	0.5	497	138	BE623310	BE623310 ut98h10.x
492	402	BF233441	BF233441 602024228	565	27	0.5	500	107	AU087396	AU087396 AU087396
493	402	BE039664	BE039664 OC02B10.O	566	27	0.5	500	224	AQ135902	AQ135902 HS_3060_A
494	402	BF233441	BF233441 yu27c10.r1	567	27	0.5	501	227	AQ373119	AQ373119 RPCI11-14
495	403	BF5140	BF5140 y994g06.s1	568	27	0.5	501	249	AZ768417	AZ768417 lM0568H22
496	403	AA213318	AA213318 RPCI-11-4	569	27	0.5	502	243	AZ422665	AZ422665 lM0201M20
497	405	AA733917	AA733917 AV733917	570	27	0.5	505	118	AW632884	AW632884 b101d08.x
498	406	AA620163	AA620163 vo63c08.r	571	27	0.5	507	16	A1155607	A1155607 ue02d03.r
499	406	AA2414121	AA2414121 HS_3074_A	572	27	0.5	507	229	AQ459895	AQ459895 HS_5131_B
500	409	AA474852	AA474852 vo53g10.s	573	27	0.5	508	248	AZ744632	AZ744632 RPCI-24-1
501	409	BE625776	BE625776 ut98h10.y	574	27	0.5	508	251	AZ871469	AZ871469 2M018A23
502	409	BF687773	BF687773 602066822	575	27	0.5	509	251	AZ670981	AZ670981 wb12d12.x
503	411	AA454820	AA454820 zx79b07.s	576	27	0.5	513	24	A1742081	A1742081 wq38g03.x
504	411	AA1493846	AA1493846 qz51f04.x	577	27	0.5	513	106	AT005187	AT005187 AT005187
505	413	AA454809	AA454809 zx79a04.s	578	27	0.5	514	19	A1399632	A1399632 NCSC5B11T

287	28	0.5	455	23	AI689088	AI689088 tx81f08.x	c 360	27	0.5	182	110	AV742759	AV742759 AV742759
288	28	0.5	456	22	AI630179	AI630179 ae26a08.r	c 361	27	0.5	191	8	AA501754	AA501754 ng05a02.s
289	28	0.5	457	16	AI132403	AI132403 uo24c05.x	c 362	27	0.5	191	18	AI251979	AI251979 qv57g09.x
290	28	0.5	476	22	AI601747	AI601747 fc04c01.x	c 363	27	0.5	191	257	T02709	T02709 0064M3 gmbp
291	28	0.5	476	107	AU096057	AU096057 AU096057	c 364	27	0.5	192	18	AI305591	AI305591 qw47b08.x
292	28	0.5	477	19	AI342754	AI342754 qo26602.x	c 365	27	0.5	193	18	AI308401	AI308401 tb43e01.x
293	28	0.5	477	188	R99446	R99446 yq79g03.r1	c 366	27	0.5	193	249	AZ767740	AZ767740 IM0567N09
294	28	0.5	477	238	AZ090343	AZ090343 RPCI-23-4	c 367	27	0.5	196	4	AA250310	AA250310 mw60a12.r
295	28	0.5	483	248	AZ729776	AZ729776 RPCI-24-9	c 368	27	0.5	200	3	AA208868	AA208868 mw75g02.r
296	28	0.5	496	105	AL370886	AL370886 MCB4A0806	c 369	27	0.5	204	157	D79598	D79598 HUM288B04B
297	28	0.5	515	175	BG249676	BG249676 602319788	c 370	27	0.5	207	7	AA405488	AA405488 zw39e04.s
298	28	0.5	528	244	AZ492998	AZ492998 IM0327119	c 371	27	0.5	209	20	AI422359	AI422359 tf65h10.x
299	28	0.5	529	105	AL501036	AL501036 AL501036	c 372	27	0.5	210	161	BB591008	BB591008 BB591008
300	28	0.5	537	156	CB3907	CB3907 CB3907 Dict	c 373	27	0.5	214	110	AN021716	AN021716 df27d01.y
301	28	0.5	546	240	AZ259147	AZ259147 RPCI-23-1	c 374	27	0.5	215	163	BE076467	BE076467 CM0-BT059
302	28	0.5	547	118	AW635587	AW635587 b135f02.w	c 375	27	0.5	221	13	AA912444	AA912444 om52a02.s
303	28	0.5	564	231	AQ634217	AQ634217 RPCI-11-4	c 376	27	0.5	221	17	AI223756	AI223756 qx32d05.x
304	28	0.5	564	244	AZ456129	AZ456129 IM0258R17	c 377	27	0.5	225	161	BB590643	BB590643 BB590643
305	28	0.5	569	19	AI392580	AI392580 NCSMLF7T3	c 378	27	0.5	229	21	AI538430	AI538430 t06b05.x
306	28	0.5	575	256	B77506	B77506 T26H20TR TA	c 379	27	0.5	235	246	AZ618385	AZ618385 IM0450001
307	28	0.5	584	107	AU029437	AU029437 AU029437	c 380	27	0.5	236	27	AV239760	AV239760 AV239760
308	28	0.5	606	244	AZ448954	AZ448954 IM0246F22	c 381	27	0.5	237	106	AL588277	AL588277 AL588277
309	28	0.5	607	11	AA769588	AA769588 n242b06.s	c 382	27	0.5	238	4	AA238681	AA238681 mx82h09.r
310	28	0.5	612	249	AZ770203	AZ770203 IM0571E10	c 383	27	0.5	238	18	AI281869	AI281869 qf68g04.x
311	28	0.5	613	123	AW982845	AW982845 HVSMEG000	c 384	27	0.5	242	127	BB151720	BB151720 BB151720
312	28	0.5	626	241	AZ280050	AZ280050 RPCI-23-1	c 385	27	0.5	242	161	BB590018	BB590018 BB590018
313	28	0.5	627	247	AZ655154	AZ655154 IM0530C02	c 386	27	0.5	246	115	AW396453	AW396453 sb03h07.y
314	28	0.5	659	153	BG435803	BG435803 602508422	c 387	27	0.5	247	17	AI203372	AI203372 qf30a05.x
315	28	0.5	681	243	AZ436700	AZ436700 IM0224K05	c 388	27	0.5	250	125	BB076877	BB076877 BB076877
316	28	0.5	695	243	AZ420291	AZ420291 IM0198M06	c 389	27	0.5	251	123	AW996630	AW996630 MRO-BN007
317	28	0.5	745	145	BF180502	BF180502 601805294	c 390	27	0.5	252	4	AA245675	AA245675 mx05g11.r
318	28	0.5	749	168	BF699770	BF699770 602127207	c 391	27	0.5	252	148	BF448500	BF448500 7n71e10.x
319	28	0.5	782	153	BG434777	BG434777 602507988	c 392	27	0.5	253	27	AV260966	AV260966 AV260966
320	28	0.5	862	172	BG025091	BG025091 602276440	c 393	27	0.5	253	27	AV260966	AV260966 AV260966
321	28	0.5	877	168	BF675571	BF675571 602083459	c 394	27	0.5	260	161	BB579507	BB579507 BB579507
322	28	0.5	897	168	BF673321	BF673321 602136175	c 395	27	0.5	261	189	TG0887	TG0887 yb72e04.r1
323	28	0.5	967	150	BF577032	BF577032 602134491	c 396	27	0.5	264	19	AI376483	AI376483 tb76f06.x
324	28	0.5	992	219	CNS0026W	AI062230 602249661	c 397	27	0.5	264	111	AW069860	AW069860 cr52e10.x
325	28	0.5	1046	169	BF790409	BF790409 602249661	c 398	27	0.5	266	18	AI311987	AI311987 ta89f11.x
326	28	0.5	1129	220	CNS02ATF	AI188988 Tetraodon	c 399	27	0.5	266	22	AI611438	AI611438 tt61b03.x
327	28	0.5	1224	142	BF962061	BF962061 601655121	c 400	27	0.5	267	19	AI313286	AI313286 ta92b04.x
328	28	0.5	1287	172	BF980431	BF980431 602304159	c 401	27	0.5	275	19	AI335410	AI335410 tb66b01.x
329	28	0.5	50	118	AW632886	AW632886 b101d10.x	c 402	27	0.5	277	19	AI345286	AI345286 tb68b11.x
330	27	0.5	54	244	AZ482068	AZ482068 IM0306M23	c 403	27	0.5	279	18	AI312292	AI312292 ta89a12.x
331	27	0.5	67	9	AA587263	AA587263 nn70f08.s	c 404	27	0.5	279	19	AI345434	AI345434 tb69c04.x
332	27	0.5	73	146	BF213384	BF213384 601845128	c 405	27	0.5	283	23	AI348772	AI348772 ta85a12.x
333	27	0.5	89	18	AI306886	AI306886 qw75f12.x	c 406	27	0.5	283	23	AI363672	AI363672 ts57f05.x
334	27	0.5	97	159	NS1554	NS1554 yf86a02.s1	c 407	27	0.5	283	124	BB067718	BB067718 BB067718
335	27	0.5	104	4	AA270996	AA270996 va85c07.r	c 408	27	0.5	284	2	AA125017	AA125017 mq73d07.r
336	27	0.5	105	249	AZ797219	AZ797219 2M0053015	c 409	27	0.5	284	131	BB340235	BB340235 BB340235
337	27	0.5	116	22	AI580620	AI580620 tb72a11.x	c 410	27	0.5	285	3	AA209063	AA209063 mw75h12.r
338	27	0.5	118	18	AI275346	AI275346 qw66c05.x	c 411	27	0.5	286	9	AA551610	AA551610 nj37b05.s
339	27	0.5	121	9	AA624364	AA624364 vn02f02.r	c 412	27	0.5	286	161	BB580952	BB580952 BB580952
340	27	0.5	124	18	AI308444	AI308444 tb43h10.x	c 413	27	0.5	287	22	AI613570	AI613570 tt80d08.x
341	27	0.5	132	18	AI308274	AI308274 tb42c04.x	c 414	27	0.5	288	27	AV259060	AV259060 AV259060
342	27	0.5	135	8	AA503588	AA503588 ng07f10.s	c 415	27	0.5	290	19	AI371951	AI371951 tb76b10.x
343	27	0.5	134	20	AI453746	AI453746 tj39g09.x	c 416	27	0.5	291	156	C22126	C22126 C22126 Miya
344	27	0.5	137	103	AI910910	AI910910 tq39f11.x	c 417	27	0.5	293	131	BB346822	BB346822 BB346822
345	27	0.5	139	122	AW885758	AW885758 rc4-OT007	c 418	27	0.5	298	12	AA853343	AA853343 NHTBcae05
346	27	0.5	143	18	AI275333	AI275333 qw66c03.x	c 419	27	0.5	298	187	R60340	R60340 yho4f03.s1
347	27	0.5	143	18	AI305997	AI305997 qw74a07.x	c 420	27	0.5	299	15	AI021015	AI021015 ua98f12.r
348	27	0.5	144	18	AI306026	AI306026 qw74a03.x	c 421	27	0.5	299	127	BB161124	BB161124 BB161124
349	27	0.5	143	18	AI312962	AI312962 ta80b01.x	c 422	27	0.5	300	149	BF459325	BF459325 UI-M-B21-
350	27	0.5	147	137	BE553062	BE553062 946088H04	c 423	27	0.5	300	230	AO582335	AO582335 RPCI-11-4
351	27	0.5	149	18	AI252259	AI252259 qv24c04.x	c 424	27	0.5	302	114	AA305242	AA305242 xr81b12.x
352	27	0.5	152	112	AW164498	AW164498 se73e01.y	c 425	27	0.5	301	122	AW960326	AW960326 EST372337
353	27	0.5	153	131	BB350763	BB350763 BB350763	c 426	27	0.5	302	161	BB585511	BB585511 BB585511
354	27	0.5	156	19	AI350028	AI350028 tb63g01.x	c 427	27	0.5	308	137	BE588294	BE588294 196942 BA
355	27	0.5	162	171	BF965460	BF965460 602276904	c 428	27	0.5	309	4	AA271459	AA271459 va83d07.r
356	27	0.5	166	170	BF886776	BF886776 PM4-TN017	c 429	27	0.5	309	148	BF440552	BF440552 BS2900021
357	27	0.5	169	118	AW633013	AW633013 b102h01.x	c 430	27	0.5	309	160	BB547267	BB547267 BB547267
358	27	0.5	171	17	AI246069	AI246069 qv52f03.x	c 431	27	0.5	312	143	BF016061	BF016061 uy30c05.y
359	27	0.5	178	28	AV313767	AV313767 AV313767	c 432	27	0.5	314	23	AI651938	AI651938 w51c12.x

SUMMARIES									
Result	No.	Score	Query	Match	Length	DB	ID	Description	
	1	631	12.0	791	149	BF509252		BF509252 UI-H-B14-	87
	2	596	11.3	596	136	BE504550		BE504550 hz58a05.x	77
	3	576	10.2	734	175	BG256731		BG256731 602370939	76
	4	536	10.2	648	122	AN973437		AN973437 ES3385430	76
	5	511	9.7	588	118	AN629426		AN629426 h137b12.x	71
	6	472	9.0	535	103	A1911784		A1911784 w98f06.x	72
	7	460	8.7	480	23	A1655429		A1655429 ts98f06.x	73
	8	457	8.7	540	114	BF002153		BF002153 72h3a12.x	64
	9	453	8.6	542	143	AW303630		AW303630 xv21e03.x	60
	10	439	8.3	848	141	BE873840		BE873840 601483992	236
	11	431	8.2	738	122	AA282190		AA282190 EST388259	136
	12	426	8.1	470	4	AA282190		AA282190 EST388259	136
	13	423	8.0	423	23	A1680124		A1680124 tw65b01.x	136
	14	422	8.0	578	120	AW770571		AW770571 h186f11.x	136
	15	418	7.9	469	9	AA563884		AA563884 nk19e08.s	136
	16	406	7.7	461	15	A1016896		A1016896 cu31c08.x	136
	17	402	7.6	409	9	AA625961		AA625961 zu92e08.s	136
	18	384	7.3	486	23	A1681422		A1681422 tx46e01.x	136
	19	384	7.3	490	149	BF511856		BF511856 UI-H-B14-	136
	20	381	7.2	433	18	AI275866		AI275866 ql68c06.x	136
	21	381	7.2	525	144	BF111072		BF111072 7n43b02.x	136
	22	368	7.0	390	148	BF464410		BF464410 7p36h09.x	136
	23	359	6.8	888	155	BE571128		BE571128 602591748	136
	24	354	6.7	391	10	AA630616		AA630616 ac11e07.s	136
	25	324	6.1	738	32	AV710329		AV710329 AV710229	136
	26	313	5.9	729	153	BG432960		BG432960 602495783	136
	27	311	5.9	406	18	AI299646		AI299646 gn12a10.x	136
	28	311	5.9	410	14	AA987361		AA987361 cr92q04.s	136
	29	305	5.8	356	13	AA911125		AA911125 ce69a12.s	136
	30	296	5.6	405	16	AA140790		AA140790 co15a04.x	136
	31	289	5.6	442	7	AA447433		AA447433 zw93f10.r	136
	32	288	5.5	505	143	BF058642		BF058642 7k33b03.x	136
	33	278	5.3	357	164	BE218464		BE218464 hv38h11.x	136
	34	269	5.1	802	153	BG387646		BG387646 602412496	136
	35	240	4.6	257	12	AA854212		AA854212 aj72a03.s	136
	36	237	4.5	535	137	BE551003		BE551003 7b67b03.x	136
	37	231	4.4	446	158	H93424		H93424 ys76a11.s1	136
	38	222	4.2	464	146	BF222521		BF222521 7p53q08.x	136
	39	220	4.2	271	5	AA333118		AA333118 EST37155	136
	40	220	4.2	310	111	AW089775		AW089775 xd08a10.x	136
	41	218	4.1	318	144	BF094231		BF094231 CMI-UT004	136
	42	205	3.9	371	158	H45693		H45693 yp23e07.s1	136
	43	198	3.8	249	11	AA747568		AA747568 nx92c01.s	136
	44	183	3.5	186	22	AI572349		AI572349 te40c12.x	136
	45	181	3.4	421	10	AA687846		AA687846 pv09a11.s	136
	46	164	3.1	306	16	AI140791		AI140791 co15a05.x	136
	47	164	3.1	357	157	F12334		F12334 HSC38h11.n	136
	48	152	2.9	606	155	BG546853		BG546853 602574035	136
	49	151	2.9	608	32	AV685306		AV685306 AV685306	136
	50	147	2.8	107	105	AL121425		AL121425 BK2f2p762H	136
	51	145	2.8	690	187	N91467		N91467 zb40d01.s1	136
	52	137	2.6	696	233	A0779594		A0779594 HS_5567_B	136
	53	131	2.5	404	156	D20082		D20082 HUMGS01054	136
	54	127	2.4	257	6	AA376799		AA376799 EST89290	136
	55	127	2.4	496	7	AA447434		AA447434 zw93f10.s	136
	56	125	2.4	176	174	BF150040		BF150040 nad44a02.	136
	57	120	2.3	408	11	AA755902		AA755902 vu04g01.r	136
	58	117	2.2	639	190	W20070		W20070 zb40d01.r1	136
	59	112	2.1	211	159	N93765		N93765 KR2745F Hum	136
	60	108	2.0	440	158	H44781		H44781 yp20g09.r1	136
	61	101	1.9	725	144	BF104916		BF104916 601822648	136
	62	100	1.9	683	136	BE532986		BE532986 601235013	136
	63	98	1.9	323	125	BB094813		BB094813 BB094813	136
	64	98	1.9	529	257	B81200		B81200 CIT-HSP-202	136
	65	97	1.8	219	139	BE706461		BE706461 RCI-HT026	136
	66	90	1.7	137	175	BG271682		BG271682 na152a11.	136
	67	87	1.7	388	118	AW578720		AW578720 RCI-CT026	136

87	1.7	437	8	AA511220	
77	1.5	429	145	BF189753	
76	1.4	193	122	AW948932	
76	1.4	314	10	AA687776	
71	1.3	741	141	BE891040	
64	1.2	602	230	AQ539049	
60	1.1	236	144	BF085557	
60	1.1	380	168	BF713255	
56	1.1	944	145	BF142489	
51	1.0	447	189	T64867	
50	0.9	372	22	AI604717	
50	0.9	462	7	AA399798	
50	0.9	560	138	BE634375	
49	0.9	361	164	BE150448	
49	0.9	421	23	AI645606	
46	0.9	296	133	BB414039	
46	0.9	316	158	H45741	
46	0.9	341	32	AV648328	
44	0.8	208	161	BB589397	
44	0.8	421	13	AA899586	
44	0.8	434	116	AW483768	
44	0.8	455	14	AI012649	
43	0.8	347	122	AW973436	
43	0.8	751	251	AZ902024	
43	0.8	818	175	BG296676	
42	0.8	486	13	AA939593	
42	0.8	504	16	AI120127	
42	0.8	595	143	BF043498	
41	0.8	324	243	AZ424160	
41	0.8	385	110	AW048929	
41	0.8	445	142	BE948523	
41	0.8	475	138	BE653757	
41	0.8	484	173	BG090245	
41	0.8	608	149	BF471938	
40	0.8	338	146	BF282055	
40	0.8	344	16	AI120130	
40	0.8	355	102	AI842943	
40	0.8	387	23	AI642950	
40	0.8	442	15	AI060512	
40	0.8	456	236	AQ855541	
40	0.8	460	237	AZ019111	
40	0.8	482	138	BE691459	
40	0.8	493	113	AW212954	
40	0.8	498	3	AA177347	
40	0.8	499	141	BE852737	
40	0.8	509	169	BF744094	
40	0.8	539	174	BG143529	
40	0.8	540	174	BG143657	
40	0.8	567	229	AQ485108	
39	0.7	304	4	AA254064	
39	0.7	305	128	BB201616	
39	0.7	327	128	BB205104	
39	0.7	403	141	BE847511	
39	0.7	459	3	AA154932	
39	0.7	493	168	BF714087	
39	0.7	540	130	BB284283	
39	0.7	583	17	AI172172	
38	0.7	322	128	BB198397	
38	0.7	329	117	AW531331	
38	0.7	580	149	BF451878	
37	0.7	526	244	AZ448028	
37	0.7	878	136	BE540142	
35	0.7	416	163	BE104755	
34	0.6	278	133	BB423438	
34	0.6	648	137	BE570606	
33	0.6	555	227	AQ323593	
33	0.6	670	105	AL036285	
32	0.6	269	134	BB474715	
32	0.6	333	156	C99689	
31	0.6	392	175	BG234054	
30	0.6	382	155	BG555674	
30	0.6	322	13	AA887581	
30	0.6	417	241	AZ330514	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:34:47 : Search time 3712.66 Seconds
(without alignments)
13420.564 Million cell updates/sec

Title: US-09-512-581-1
Perfect score: 5271
Sequence: 1 ccggagagcccgagtgag.....gtaaaaaaaaaaaaaaaa 5271

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est33:*
111: gb_est34:*
112: gb_est35:*
113: gb_est36:*
114: gb_est37:*
115: gb_est38:*
116: gb_est39:*

THIS PAGE BLANK (USPTO)

Db 736
736 wwwwww 743

RESULT 15

AAF58259

ID AAF58259 standard; DNA; 936 BP.

XX AC

XX AAF58259;

XX 24-APR-2001 (first entry)

XX DE

XX Oligonucleotide D2004.

XX KW

XX Electron-transfer group; ETM; mismatch; genotyping;

XX KW

XX gene expression; ss.

XX OS

XX Synthetic.

XX PN

XX WO200107665-A2.

XX XX

XX 01-FEB-2001.

XX PD

XX 26-JUL-2000; 2000WO-US20476.

XX PF

XX 26-JUL-1999; 99US-0145695.

XX PR

XX 17-MAR-2000; 2000US-0190259.

XX XX

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX PA

XX Umek RM;

XX PI

XX WPI; 2001-159728/16.

XX DR

XX Nucleic acids containing electron-transfer group, useful as labels in

XX PT

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX PT

XX a single surface

XX XX

XX Example 6; Page 128; 159pp; English.

XX PS

XX The present invention relates to a composition comprising two nucleic

XX CC

XX acids each containing an electron-transfer group (ETM) having

XX CC

XX different redox potentials. The invention is used for electronic

XX CC

XX detection of nucleic acids, especially of substitutions (mismatches)

XX CC

XX and single-nucleotide polymorphisms, e.g. for genotyping,

XX CC

XX monitoring gene expression.

XX XX

XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 3.1%; Score 165.6; DB 22; Length 936;

Best Local Similarity 0.8%; Pred. No. 1.6e-25;

Matches 6; Conservative 494; Mismatches 228; Indels 0; Gaps 0;

Qy 4544 atacatttggtcacatgcttttagccatacacatggaacattgactatggagcttgt 4603

Db 16 75

Qy 4604 gaaagtgaatgtagcgtatgtagacataaagaacattgtataatcttttt 4663

Db 76 135

Qy 4664 ctttttttaagtcttctgattctctgaagtgctgtatagctttttatctcggttttaa 4723

Db 136 195

Qy 4724 ctgacagaccacgactgtttattggtatctattgtattgaaagaattgttagatagat 4783

Db 196 255

Qy 4784 cttaagcagtaactgtcagtgctgtattgtatttctgtcaatttttactgtgaaaaa 4843

Db 256 315

Qy 4844 aattgttttcaacaattggtgctcattttcttgatgctcactattgttgagagattaaat 4903
Db 316 375
Qy 4904 ggtctcttccctttgtgtatcttactagtgtttactcctcctgggaccccttaattctcaga 4963
Db 376 435
Qy 4964 ggtgctaaattgtctgccattacaccagaagatgcctctgatagaggagacaaccatgca 5023
Db 436 495
Qy 5024 aattgtgaaatagctcctgaagtctcttgattactttacacctcagttattgttgcacca 5083
Db 496 555
Qy 5084 gaattttctgcccctttcatgccaatgaaaatttttaagaagaagatttaagatttttaa 5143
Db 556 615
Qy 5144 tttaaaagagtgtgtataaaaataatgtactgaattctttatcccatctttatcatccttt 5203
Db 616 675
Qy 5204 cagtttttataatctactgtatcaataaaattctgttaatttgatgagtaaaaaaaa 5263
Db 676 735
Qy 5264 aaaaaaaa 5271
Db 736 743

Search completed: September 25, 2001, 19:52:23

Job time: 19461 sec


```
Db 76 135
Qy 4664 4723
Db 136 195
Qy 4724 4783
Db 196 255
Qy 4784 4843
Db 256 315
Qy 4844 4903
Db 316 375
Qy 4904 4963
Db 376 435
Qy 4964 5023
Db 436 495
Qy 5024 5083
Db 496 555
Qy 5084 5143
Db 556 615
Qy 5144 5203
Db 616 675
Qy 5204 5263
Db 676 735
Qy 5264 5271
735

RESULT 14
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
FA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
```

```
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 3.1%; Score 165.6; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.6e-25;
Matches 6; Conservative 494; Mismatches 228; Indels 0; Gaps 0;

Qy 4544 atacatttggctcacatgcttttagccacacacatggttaacattgactgagcttgt 4603
Db 16 75
Qy 4604 gaaagtgtatgtgcgctgtatgtagacataaagaagaactgttaaatcttttt 4663
Db 76 135
Qy 4664 ctttttttaagtctgtgatttctgaagctgtgtatgtatttctggtctttaa 4723
Db 136 195
Qy 4724 ctgacagtaccgactgtttatgtgactctattgttgaagaactttgttagtagatag 4783
Db 196 255
Qy 4784 cttaagcagtaactgtgcagtggttattgtattttctgtaatttctactgtgaaaaa 4843
Db 256 315
Qy 4844 aatttgttttcaacaataattgggtcacttttctgtgactcactatttctgtggagagtaaat 4903
Db 316 375
Qy 4904 ggtctcttcccttgtgtatcttaactagtgtttactctctggcacccttaattctcaga 4963
Db 376 435
Qy 4964 ggtgtataattgtctgcccattacacacagaagatgcctctgatagagacacacatgca 5023
Db 436 495
Qy 5024 aattgtgaaatagctcctgaagtcttcttgattacttcttaacactcagattgattgtccca 5083
Db 496 555
Qy 5084 gaatttctgaccttctcatgccaatgaaattttaagaagaagattttaagatttttaa 5143
Db 556 615
Qy 5144 tttaaaagagtgtgtataaaataatgactgaattcttccatccatttatccatcttt 5203
Db 616 675
Qy 5204 cagttttttattactactactgtatcaataaattctgttaattgtaagtgaataaaaaaa 5263
Db 676 735
Qy 5264 aaaaaaa 5271
```



```
OS Synthetic.
XX WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000NO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX Example 6; Page 127; 159pp; English.
XX PS The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 3.1%; Score 165.6; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.6e-25;
Matches 6; Conservative 494; Mismatches 228; Indels 0; Gaps 0;

QY 4544 atacatttggtgcagtgctgtatgacataaagaacattgtataatctttttt 4603
Db 16 www. ....
QY 4604 gaaagtgaatgctgctgtatgacataaagaacattgtataatctttttt 4663
Db 76 www. ....
QY 4664 ctttttttaattgtctgctgaagtctgtatagcttttttatctgcgctttaa 4723
Db 136 www. ....
QY 4724 ctgacagtcaccgactgttttattggtatctattgtattgaaagaattgttaggatagat 4783
Db 196 www. ....
QY 4784 cttaagcagtaactgtcagttgtattgtattgtattttctgtcaatttttactgtgaaaaa 4843
Db 256 www. ....
QY 4844 aattgttttcaacaattggttcattttctgtatgctcaactattgttgagagattaaat 4903
Db 316 www. ....
QY 4904 ggtctctcccttgtgtatcttaccatgtgttttactccttggtgcaccccttaattctcaga 4963
Db 376 www. ....
QY 4964 ggtgctaaattgtctgccattacacagaaggatgctctgtatgaggagacacaccatgca 5023
Db 436 www. ....
QY 5024 aattgtgaaatagctctgaagtcttctggtatcttaccactcagttattgttgcacca 5083
Db 496 www. ....
```

```
QY 5084 gaatttttgccttttccatggcaatgaaaaattttaagaagaagatttaagatttttaa 5143
Db 556 www. ....
QY 5144 tttaaagagtggttataaaaaataatgactgaattcttttcccatctttatcatccttt 5203
Db 616 www. ....
QY 5204 cagtttttatactactgtatcaataaaattctgttaatttgaatgagtaaaaaaaa 5263
Db 676 www. ....
QY 5264 aaaaaaaa 5271
Db 736 www. ....

RESULT 13
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX AC AAF58254;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1875.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000NO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface -
XX
XX Example 6; Page 127; 159pp; English.
XX PS The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 3.1%; Score 165.6; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.6e-25;
Matches 6; Conservative 494; Mismatches 228; Indels 0; Gaps 0;

QY 4544 atacatttggtgcagtgctgtatgacatacactgtaacattgactatgaggtcttgt 4603
Db 16 www. ....
QY 4604 gaaagtgaatgctgctgtatgacataaagaacattgtataatctttttt 4663
Db 76 www. ....
```



```

RESULT 9
AAZ14459
ID AAZ14459 standard; cDNA; 284 BP.
XX
AC AAZ14459;
XX
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:1928.
XX
XX Human: gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
XX
XX WO9938972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX
XX 28-JAN-1998; 98US-0072910.
XX
XX 24-FEB-1998; 98US-0075954.
XX
XX 31-MAR-1998; 98US-0080114.
XX
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
XX Claim 1; Page 1096; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
XX Sequence 284 BP; 89 A; 43 C; 46 G; 103 T; 3 other;
XX
XX Query Match 4.9%; Score 258.8; DB 20; Length 284;
XX Best Local Similarity 97.8%; Pred. No. 3.3e-45;
XX Matches 272; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

```

```

QY 4970 aaattgtctccattacaccaggaaggatgcctctgatagaggacacaccatgcaaatgt 5029
Db 1 aaattgtctccattacaccaggaaggatgcctctgatagaggacacaccatgcaaatgt 60
QY 5030 gaaatagctcctgaagttctctggattactttacacctcagttattgtccccaattt 5089
Db 61 gaaatagctcctgaagttctctggattactttacacctcagttattgtccccaattt 120
QY 5090 tcggcccttcctggaatgaaatatttaagaagaagaagatttaaaagtattttaatttaa 5149
Db 121 tcggcccttcctggaatgaaatatttaagaagaagaagatttaaaagtattttaatttaa 180
QY 5150 agagtggttataaaataatgtaactgttatttcccatctttatcatccttccagttt 5209
Db 181 agagtggttataaaataatgtaactgttatttcccatctttatcatccttccagttt 240
QY 5210 ttataatctactgtatcaataaaattctgtaatttga 5247
Db 241 ttataatctactgtatca--taaatctgtaantnga 276

```

```

RESULT 10
AAF18338
ID AAF18338 standard; DNA; 1217 BP.
XX
AC AAF18338;
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polynucleotide sequence SEQ ID 357.
XX
XX Human: lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulneryary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
XX
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI: 2000-587514/55.
XX P-PSDB; AAB58462.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -
XX
XX Claim 1; Page 815; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX associated proteins and polynucleotide sequences, their agonists, and
XX antagonists may have neuroprotective; cytostatic; cardioactive;
XX immunomodulatory; muscular active general; vulneryary; gastrointestinal
XX general; nephrotropic; antiinfective; gynecological; or antibacterial
XX activity. The invention also includes antibodies specific for the
XX protein or polynucleotide sequences. The lung cancer associated
XX polynucleotide sequences may be used for detection of lung cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The proteins may be used to treat
XX disorders such as neural, immune, muscular, reproductive,

```


PR 28-APR-1998; 98US-0069047.
PA (GEST) GENSET.
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-038446/03.
DR P-PSDB; AAY65247.
XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 1; Page 495; 837pp; English.
XX
CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AA242265 to
CC AA243052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 351 BP; 107 A; 63 C; 80 G; 99 T; 2 other;

Query Match 6.3%; Score 333; DB 21; Length 351;
Best Local Similarity 98.6%; Pred. No. 1e-60; Indels 1; Gaps 1;
Matches 344; Conservative 2; Mismatches 2;
QY 19 agcggagtagcagtcgcaaccggagggtagaataatttctgtcatggtctattcaa 78
Db 1 agcggagtagcagtcgcaaccggagggtagaataatttctgtcatggtctattcaa 59
QY 79 agactaggaccatgatggaaataatacatatccgcctggggtcagaagaatacatgata 138
Db 60 agactaggaccatgatggaaataatacatatccgcctggggtcagaagaatacatgata 119
QY 139 aaatatcaagaggagatggtgagacgattaaagatggttgaaactttttatggata 198
Db 120 aaatatcaagaggagatggtgagacgattaaagatggttgaaactttttatggata 179
QY 199 tggaccagactctgaagagaaaggagctttatttaaacctagctttacattctgctt 258
Db 180 tggaccagactctgaagagaaaggagctttatttaaacctagctttacattctgctt 239
QY 259 cagatttttttcaagcatcctggttaagatggttcgcttactgtagctgctgcttg 318
Db 240 cagatttttttcaagcatcctggttaagatggttcgcttactgtagctgctgcttg 299
QY 319 ctgatattttcaggtattatgctctgaagctccctacacatccctga 367
Db 300 ctgatattttcaggtattatgctctgaagctccctacacatccctga 348

RESULT 8
ID AAC02817
XX AAC02817 standard; cDNA; 295 BP.
AC AAC02817;
XX

DT 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 2815.
DE
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
OS
XX
PW EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR P-PSDB; AAG02811.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 2815; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

Query Match 5.5%; Score 292; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 3.7e-52;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2562 gtctgatggctacttggaaatgaaataataatcacagtaaatcagggaacttctaccttaaga 2621
Db 1 gtctgatggctacttggaaatgaaataataatcacagtaaatcagggaacttctaccttaaga 60
QY 2622 ttgtcaacaacaataattgcattgtagagactgagacaggaaggggaaataatgataa 2681
Db 61 ttgtcaacaacaataattgcattgtagagactgagacaggaaggggaaataatgataa 120
QY 2682 ccagatatgtcagctctgagacttgcgtggagtgctattgtgaagctggcacaagaa 2741
Db 121 ccagatatgtcagctctgagacttgcgtggagtgctattgtgaagctggcacaagaa 180
QY 2742 cccgtgtaccatgaaatcatcattagaaataatcagctatgtgcatagctatcaac 2801
Db 181 cccgtgtaccatgaaatcatcattagaaataatcagctatgtgcatagctatcaac 240
QY 2802 gatgaatgtcatcaagtaagacaagtggttgcgcagaaacttcacaagccc 2853
Db 241 gatgaatgtcatcaagtaagacaagtggttgcgcagaaacttcacaagccc 292

Stache-Crain B, Sudguth-Klinger J, Williams LT;
WPI; 1999-494092/41.

differentially expressed in different cell types

Claim 1: Page 1324: 2479pp; English.

The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ17779. Also, the present invention describes a method of detecting differentially expressed genes corresponding to the sequences given in AAZ12532 to AAZ17779.

differentially expressed gene product in a test sample from a cancerous state or a noncancerous state, comprising detecting the expression of the gene product in the test sample and comparing the expression of the gene product in the test sample with the expression of the gene product in a control sample, wherein the gene product is suspected of being cancerous, where the gene product is one of the 5248 polynucleotide sequences given in AM12532 to

be used for a variety of purpose, e.g. detection of expression, mapping, tissue typing or profiling, forensics, genetic analysis, detection of polymorphisms. Polypeptides encoded by the poly-

therapeutic purposes. The polynucleotides may also be used in arrays for diagnostics (which may be used to determine function of encoded protein) and to detect differences in expression between cells (e.g., to identify abnormal or diseased tissue).

identify a genetic predisposition or susceptibility to a disease (e.g., cancer). The polynucleotides of the invention are especially useful to identify a genetic predisposition or susceptibility to a disease (e.g., cancer). The polynucleotides of the invention are especially useful for the diagnosis, prognosis and management of colorectal cancer, lung cancer, and prostate cancer. The polynucleotides can also be used to screen for the presence of a mutation in a gene associated with colorectal cancer, lung cancer, and prostate cancer.

peptide analogues and antagonists.
Sequence 738 BP; 268 A; 145 C; 168 G; 150 T; 7 other;

```

every Match          7.8%; Score 409.4; DB 20; Length
1st Local Similarity 83.0%; Pred. No. 1.3e-76;
atches 563; Conservative 0; Mismatches 77; Indels

```

3890 agaaatgaagatgaacagaaatagtcgcccaaaaaagggttaaaagagagccgac
|||||
51 aaaaatgaagatgaacagaaatagtcgcccaaaaaagggttaaaagagagccgac

3950 acctcttggtgaggtacaccaaaagaagagccacaatgaaaacttctaa
111 acctcttggtgaggtacaccaaaagaagagccacaatgaaaacttctaa

4010 caaaaaaaatctggacctccagcaccagaggaggaggagaagaagaagaagac
|||
171 caaaaaaaatctggacctccagcaccagaggaggaggagaagaagaagaagac

4070 aaatcagggaacagaagtccaaaagcaaacagaccgaggtctcaagggaagcac
|||||
231 aaatcagggaacagaagtccaaaagcaaacagaccgaggtctcaagggaagcac

4130 agcagaatctctgaatctctagtgaatgaaatccacacagtcacacaccacagaa

described is with the least one all by one 779. The which can levels, s and nucleotides and construct of an s between man, to e such as: ed in the t cancer, for

Gaps 7:

aa 3949
||
aa 110

ag 4009
||
ag 170

gg 4069
||
gg 230

ag 4129
||
ag 290

cg 4189

QY 2007 gcaggcttgaactgcttaaggttactctcatttaccacatcccatctcatttccattctgct 2056
DB 662 tcaggactgaactcttaaggttctgcttttaccacatcccatctcatttccattctgca 721
QY 2067 gaacattgaatcatctactgcttctgctgaaatgatgatgaaagtagcagaagct 2126
DB 722 gagacataagtccttctgttacagtgcctaaagaatgaggtgacaggttagcagaagct 781
QY 2127 gcaatacaaaatctcaaaaacacaggagcaaaattgaagggattttccacacatcaga 2186
DB 782 gctattcaaaatcttagaatacacaggtccacaaaatagaacacagaccttcccacagatga 841
QY 2187 tcagcttctgcttctgcttttaccacaaaatctaaaaaagagcccccccgctcaagcaaaa 2246
DB 842 tcgaccttaattcccatcttaccacaaagcaaaaggggtactccacacaaagcaaaa 901
QY 2247 tatgccattcaattgtaccatgcgatattttctagttaaagagaccaggttgcacagata 2306
DB 902 caggctgtgcactgtatacagccatattccaaaataaagaagtcaggttgcacagatt 961
QY 2307 ttgagcctctgcataagagcctagatccaaagcaacctggaacatctcatacaccacttg 2366
DB 962 ttgagcactcagtaggtctgaatgctgtagtgcagagaacaactataactccatta 1021
QY 2367 gttactattggtcaattgcttctcctgcaacctgatcaattgctgctccttggaaatct 2426
DB 1022 gttcattggccacattctctatgttagcaccagatcagttgttcccaatgaaatct 1081
QY 2427 tgggtagctacttctcattgtgaagatctctcattgaatgatcgcttccagggaagaaag 2486
DB 1082 gtagtagcaaaatttattgtgaagatctgctaagtgaatgcaggttcaaacagggtgaagaa 1141
QY 2487 acaactaaacttgggttccagatgaagaagatctctctgagacaatgggtcaaaattcag 2546
DB 1142 aatggaaactgtgtctccagatgaagaggtttccctctgaagtactgacgaaggtacag 1201
QY 2547 gctattaaaatggttctgctgctacttggaatgaaataataatcacagtaaatcagga 2606
DB 1202 gcaattaaactctggaaggtgctgttggtgtagaaacacacagctcaaatctgccc 1261
QY 2607 acttctaccttaagattgtcaacacaaatattgcatagtgatgagagcttgacagaaacag 2666
DB 1262 aattcaacctctggttattatcagcagatgttggttagtgaggtgacctgacagagcaa 1321
QY 2667 gggaaaattagtaaacagatatgtcagctgtcgtgagacttgcctggagtgctattggt 2726
DB 1322 aagagatcagtaaatctgatatgtctcgtctgctgattagctgctggttagtgccataatg 1381
QY 2727 aagctggcacaagaacctgttaccatgaaatcaccattagacaaatcagctatgt 2786
DB 1382 aagctgtcaggaacctgttaccatgaaatattaccacagacaggttccagctctgt 1441
QY 2787 gcattagctatcaacgatgaatgctatcaagtaagacaagtgttggccagaaacttcaac 2846
DB 1442 gcactgttattaatgatggtgttaccagtaagcagagatatgttgcagaaagctgcat 1501
QY 2847 aaaggcctttcccgcttaccggcttccactgagtatatggcaatcctgtgccccttggca 2906
DB 1502 aaggcactgtgaaagtactgctcccttggagtgatatggcgatcttggccttggctgccc 1561
QY 2907 aagatcctgttaagagagagaagagctcatgctaggaatgttggtaaaaaataataat 2966
DB 1562 aaagatcctgtgaagagagagagagcacaagcagcaaatgttactgaaaaataatcagt 1621
QY 2967 gtaagggggagtagtctgaagcagcagctgctgttagtgaaaaattattgctcttcta 3026
DB 1622 atacagaggaataacattaaagcagaacctatgctactgagaataattattactactgttg 1681
QY 3027 ccagagtagtgttctccatatacaattcaacttggacatgacccagattatgtcaaa 3086
DB 1682 cctgaatgatgttccatcatatgattcacctgtgacccatgacccagatttaccaga 1741
QY 3087 gtacaggatattgaacacacttaagatgttaagaagatgttcttgggttcttctggaaata 3146

RESULT 4

AAV87629
ID AAV87629 standard; cDNA; 439 BP.

XX AAV87629;

XX AC
XX AC
XX 12-FEB-1999 (first entry)XX DE
XX DE EST clone DV17.XX KW
XX KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
XX KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
XX KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
XX KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.XX OS
XX OS Homo sapiens.XX PN
XX PN WO9845437-A2.XX PD
XX PD 15-OCT-1998.XX PF
XX PF 10-APR-1998; 98WO-US06956.XX PR
XX PR 10-APR-1997; 97US-0837312.XX PA
XX PA (GEM) GENETICS INST INC.XX PI
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;XX PI
XX PI Racie LA, Spaulding V, Treacy M;XX DR
XX DR WPI; 1999-070078/06.XX PT
XX PT New polynucleotides encoding human secreted proteins - derived from
XX PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX PT ovary, pituitary, retina and colon cDNA librariesXX PS
XX PS Claim 1; Page 126; 641pp; English.

Qy	3001	ttagtgaaaaattattgtcttctaccagatgattgttgttcatacataaattcaccttt	3069
Db	3085	ttagtgaaaaattattgtcttctaccagatgattgttgttcatacataaattcaccttt	3144
Qy	3061	tggcacatgccagattatgtccaagtacaggatatatgaacaacttaagaattttaag	3120
Db	3145	tggcacatgccagattatgtccaagtacaggatatatgaacaacttaagaattttaag	3204
Qy	3121	aatgtctttggtttgtcttgaaaattaataatggctcaaaaaTgaaaaaaacacgtcacgctt	3180
Db	3205	aatgtctttggtttgtcttgaaaattaataatggctcaaaaaTgaaaaaaacacgtcacgctt	3264
Qy	3181	ttatcagaaaagatgytgtagaaaatatataaacacaaaaagatgcocaaaggaccagatgatg	3240
Db	3265	ttatcagaaaagatgytgtagaaaattataaacacaaaaagatgcocaaaggaccagatgatg	3324
Qy	3241	caaaaaatgaatgaaaaaactgtcacctgtgtgtgtgtgtgtgtgtgtccatgaatatcatgtcaa	3300
Db	3325	caaaaaTgaatgaaaaaactgtcacctgtgtgtgtgtgtgtgtgtgtccatgaatatcatgtcaa	3384
Qy	3301	agagtactacatacagtttggaattctctaagaccccgttactaccagctcgttttcttca	3360
Db	3385	agagtactacatacagtttggaattctctaagaccccgttactaccagctcgttttcttca	3444
Qy	3361	ctcaacctgacaagaatttcagTaaacacacaaaaattattctgtcctctgaaatgaattcat	3420
Db	3445	ctcaacctgacaagaatttcagTaaacacacaaaaattattctgtcctctgaaatgaattcat	3504
Qy	3421	ttttcactctgaaaacctaataaacacaaatgtttcttaggagctgttaaagccaacttt	3480
Db	3505	ttttcactctgaaaacctaataaacacaaatgtttcttaggagctgttaaagccaacttt	3564
Qy	3481	catacgaggcaagcaattctcagacaaatcatcacgaattggaacTgtaaagcaattgcaa	3540
Db	3565	catacgaggcaagcaattctcagacaaatcatcacgaattggaacTgtaaagcaattgcaa	3624
Qy	3541	gcagcagctcaaatccaagctctctctggaagaaataaaagggagcttgatgtctgaaa	3600
Db	3625	gcagcagctcaaatccaagctctctctggaagaaataaaagggagcttgatgtctgaaa	3684
Qy	3601	Tgatacacagtgaatatgaattacacaaatgtcttccactttgccggggaaaaaaagtg	3660
Db	3685	Tgatacacagtgaatatgaattacacaaatgtcttccactttgccggggaaaaaaagtg	3744
Qy	3661	acaagagacgactctgtattTaaagctTgaattTggaagacctagaggccaggaaaa	3720
Db	3745	acaagagacgactctgtattTaaagctTgaattTggaagacctagaggccaggaaaa	3804
Qy	3721	aaacgccgttcacagacaggaggagaaattaggTatggatGacttGactaaagtTgtgtac	3780
Db	3805	aaacgccgttcacagacaggaggagaaattaggTatggatGacttGactaaagtTgtgtac	3864
Qy	3781	agaaacagaaaacctaaaggcagtacgaaagTCggaaaagaggccaTacggcttcagaatt	3840
Db	3865	agaaacagaaaacctaaaggcagtacgaaagTCggaaaagaggccaTacggcttcagaatt	3924
Qy	3841	ctgatgaacagcagtgccctgaggaagagagcctcaagasaGaTaTaTtagaaaaTgaag	3900
Db	3925	ctgatgaacagcagtgccctgaggaagagagcctcaagasaGaTaTaTtagaaaaTgaag	3984
Qy	3901	atgaacagaaatagtcgcgcaaaaaagggTaaaagagccgaccaccaaacctcttggtg	3960
Db	3985	atgaacagaaatagtcgcgcaaaaaagggTaaaagagccgaccaccaaacctcttggtg	4044
Qy	3961	gaagTaccacaaaagaagaccacaaTgaaaactTctaaaaaagggaagcaaaaaaaaaat	4020
Db	4045	gaagTaccacaaaagaagaccacaaTgaaaactTctaaaaaagggaagcaaaaaaaaaat	4104
Qy	4021	ctggacctccagcaccagagggaggaagaagaagaaGacaagTggaaatacggaac	4080
Db	4105	ctggacctccagcaccagagggaggaagaagaagaaGacaagTggaaatacggaac	4164
Qy	4081	agaagtccaaaaggcaaacgaccagTgtCaaggaGagcagacgaGagagcagaaTctc	4140

[illegible]

Db 865 aggtcttgatgctggtgggaaacatctctacgagcgtattgttcagagcatgtctttgacttaa 924
Qy 841 ttttggagctcaacaattatgatagctatttgcgtctctctgttttccccagcttgaat 900
Db 925 ttttggagctcaacaattatgatagctatttgcgtctctctgttttccccagcttgaat 984
Qy 901 ttaaattaaagagcaatgataatgaggagcgcctacaaagtgtttaaactactggtgcaaaaa 960
Db 985 ttaaattaaagagcaatgataatgaggagcgcctacaaagtgtttaaactactggtgcaaaaa 1044
Qy 961 tgtttggggcaaaagattcagaatttgcctctcacaacacagccactttggcagctgact 1020
Db 1045 tgtttggggcaaaagattcagaatttgcctctcacaacacagccactttggcagctgact 1104
Qy 1021 tgggcaggtttaaataatgataatgaggagcgcctacaaagtgtttaaactactggtgcaaaaa 1080
Db 1105 tgggcaggtttaaataatgataatgaggagcgcctacaaagtgtttaaactactggtgcaaaaa 1164
Qy 1081 attgtctcatgaaccactctgatttagcaaaagacttaacagagctatcttaaagtgggt 1140
Db 1165 attgtctcatgaaccactctgatttagcaaaagacttaacagagctatcttaaagtgggt 1224
Qy 1141 cacatgacctgaggaagctattagaatgatgtattgttgcataatgattacagctgcta 1200
Db 1225 cacatgacctgaggaagctattagaatgatgtattgttgcataatgattacagctgcta 1284
Qy 1201 aaaagattattcttctggtcaatgatacacttacttaatttggtagagagagaaacattag 1260
Db 1285 aaaagattattcttctggtcaatgatacacttacttaatttggtagagagagaaacattag 1344
Qy 1261 acaaacgatgagagtagtcgaaagaagccatgattgggacttggcccaatttataagaat 1320
Db 1345 acaaacgatgagagtagtcgaaagaagccatgattgggacttggcccaatttataagaat 1404
Qy 1321 atgctttacagtcagcagctggaaagaatgctgcacaaacagatagcattgataaagaca 1380
Db 1405 atgctttacagtcagcagctggaaagaatgctgcacaaacagatagcattgataaagaca 1464
Qy 1381 aattgctacatatattacataaataatgattgtagtcgactacttgttgaacggatct 1440
Db 1465 aattgctacatatattacataaataatgattgtagtcgactacttgttgaacggatct 1524
Qy 1441 ttgctcaatacatggttctctcacaatttagaaactacagaacgagatgaaatgcttatatt 1500
Db 1525 ttgctcaatacatggttctctcacaatttagaaactacagaacgagatgaaatgcttatatt 1584
Qy 1501 acttgtagcacaactggatttaaatgctgtgaagaatgaaatggaatgtggaaatgct 1560
Db 1585 acttgtagcacaactggatttaaatgctgtgaagaatgaaatggaatgtggaaatgct 1644
Qy 1561 aaaatctgctccgacatcaagtaagatttgccttgacttgattaaagcaaccccaaaacag 1620
Db 1645 aaaatctgctccgacatcaagtaagatttgccttgacttgattaaagcaaccccaaaacag 1704
Qy 1621 atgcccagtgccaagggcoatattttcaaaagtgtggttattacaagaataatctactgac 1680
Db 1705 atgcccagtgccaagggcoatattttcaaaagtgtggttattacaagaataatctactgac 1764
Qy 1681 ctggttaaggtccaggatttcaatgaagaatttcacacaggtgttgaagaatgattgagaaaa 1740
Db 1765 ctggttaaggtccaggatttcaatgaagaatttcacacaggtgttgaagaatgattgagaaaa 1824
Qy 1741 taagaaacagcttagaagtaacttgttagtccaacatgctcctgcgaacgagctgaaagtt 1800
Db 1825 taagaaacagcttagaagtaacttgttagtccaacatgctcctgcgaacgagctgaaagtt 1884
Qy 1801 gttgtcgtgaaataactaaagatttgggcaaccccaaaacagccctacaaatcttctctgg 1860
Db 1885 gttgtcgtgaaataactaaagatttgggcaaccccaaaacagccctacaaatcttctctgg 1944
Qy 1861 aaatgatcaagtttcttcttgagaggtatgacacctgtgcacatagatcaccaatctatca 1920
Db 1945 aaatgatcaagtttcttcttgagaggtatgacacctgtgcacatagatcaccaatctatca 2004

Qy 1921 gtgctcttattaaacaagtgaacaaatacaatagatgaacacagcagatgatgaagatgagg 1980
Db 2005 gtgctcttattaaacaagtgaacaaatacaatagatgaacacagcagatgatgaagatgagg 2064
Qy 1981 gtgtccaactgatcaagccatcagagcagctgtgaactgtttaagctactctcaattta 2040
Db 2065 gtgtccaactgatcaagccatcagagcagctgtgaactgtttaagctactctcaattta 2124
Qy 2041 cacatcccatctcatttccattctcgtcaaacatttgaatcattactggttctgctgaaaa 2100
Db 2125 cacatcccatctcatttccattctcgtcaaacatttgaatcattactggttctgctgaaaa 2184
Qy 2101 tggatgatgaaaaagttagcagaagctgcactacaaaattttcaaaaacacagaaagcaaaa 2160
Db 2185 tggatgatgaaaaagttagcagaagctgcactacaaaattttcaaaaacacagaaagcaaaa 2244
Qy 2161 ttgaagaggattttccacacatcagatcagccttgcctcctgttcttcaatcacaaaatcta 2220
Db 2245 ttgaagaggattttccacacatcagatcagccttgcctcctgttcttcaatcacaaaatcta 2304
Qy 2221 aaaagagaccccccgccaagccaaaatagccattcattgtatccatgcgataatttcta 2280
Db 2305 aaaagagaccccccgccaagccaaaatagccattcattgtatccatgcgataatttcta 2364
Qy 2281 gtaagagagaccagctttgacacagatatttggagcctctgcataaagagcctagatccaaagca 2340
Db 2365 gtaagagagaccagctttgacacagatatttggagcctctgcataaagagcctagatccaaagca 2424
Qy 2341 acctggaacatctcataaacaccattggttactattggtcattgctcctcttgcacctg 2400
Db 2425 acctggaacatctcataaacaccattggttactattggtcattgctcctcttgcacctg 2484
Qy 2401 atcaatttgcctccttggaaatcttgggttagctacttcttcttcaatgtgaaagatcttcta 2460
Db 2485 atcaatttgcctccttggaaatcttgggttagctacttcttcttcaatgtgaaagatcttcta 2544
Qy 2461 tgaatgatcggtctccaggggaaaaagacaactaaacttgggttccagatgaagaagat 2520
Db 2545 tgaatgatcggtctccaggggaaaaagacaactaaacttgggttccagatgaagaagat 2604
Qy 2521 ctctcgagacaatgggtcacaatttcaggctataaaaatgagtggttcgtacttctgaa 2580
Db 2605 ctctcgagacaatgggtcacaatttcaggctataaaaatgagtggttcgtacttctgaa 2664
Qy 2581 tgaataataatcacagtaaaatcacagaaacttcaacttaagattgcttaacaacaattctgc 2640
Db 2665 tgaataataatcacagtaaaatcacagaaacttcaacttaagattgcttaacaacaattctgc 2724
Qy 2641 atagtgtgagagacttgacagaacaggggaaaaattagtaaacccagatattgcaagctgca 2700
Db 2725 atagtgtgagagacttgacagaacaggggaaaaattagtaaacccagatattgcaagctgca 2784
Qy 2701 gacttgcctggtgagtgctatttgaagctggcaacagacccctgttaccatgaaatca 2760
Db 2785 gacttgcctggtgagtgctatttgaagctggcaacagacccctgttaccatgaaatca 2844
Qy 2761 tcacattagaacaataatcagctatgctatgctatcagctgaatgctatcaagtaa 2820
Db 2845 tcacattagaacaataatcagctatgctatgctatcagctgaatgctatcaagtaa 2904
Qy 2821 gacaagtgtttgcccagaaacttcacaaagcccttcccggtttacggcttccacttgagt 2880
Db 2905 gacaagtgtttgcccagaaacttcacaaagcccttcccggtttacggcttccacttgagt 2964
Qy 2881 atatggcaatctgtcccttctgtcacaagaatcctgttaaaagagagagatcgtatgcta 2940
Db 2965 atatggcaatctgtcccttctgtcacaagaatcctgttaaaagagagagatcgtatgcta 3024
Qy 2941 ggaatgttttggtaaaaaataaaatgtaagcggtggagatctctgaagcagcatgagctg 3000
Db 3025 ggaatgttttggtaaaaaataaaatgtaagcggtggagatctctgaagcagcatgagctg 3084

FT exon 2625..2761
FT /tag= v
FT /number= 23
FT exon 2672..2885
FT /tag= w
FT /number= 24
FT exon 2886..3090
FT /tag= x
FT /number= 25
FT exon 3091..3205
FT /tag= y
FT /number= 26
FT exon 3206..3338
FT /tag= z
FT /number= 27
FT exon 3339..3458
FT /tag= aa
FT /number= 28
FT exon 3459..3521
FT /tag= ab
FT /number= 29
FT exon 3522..3667
FT /tag= ac
FT /number= 30
FT exon 3668..3773
FT /tag= ad
FT /number= 31
FT exon 3774..4213
FT /tag= ae
FT /number= 32
FT exon 4214..4438
FT /tag= af
FT /number= 33
FT misc_signal 4314..4408
FT /tag= ag
FT /note= "Destabilizing signal"
FT exon 4439..5337
FT /tag= ah
FT /number= 34
FT polyA_signal 5312..5317
FT /tag= ai
FT misc_feature 5333..5337
FT /tag= aj
FT /note= "Poly adenylation cleavage site"
XX
XX WO200050454-A1.
XX
XX 31-AUG-2000.
XX
XX 24-FEB-2000; 2000WO-US04732.
XX
XX 24-FEB-1999; 99US-0121461.
XX
XX (TUFT) TUFTS COLLEGE.
XX
XX Soto AM, Sonnenschein C, Geck P, Szelei J;
XX
XX WPI; 2000-565451/52.
XX P-PSDB; AAY94702.
XX
XX New human androgen-induced tumor suppressor cDNA sequence termed
XX 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
XX diagnosis and treatment of prostate cancer -
XX
XX Example 4; Fig 6; 152pp; English.
XX
XX This invention relates to a human androgen-induced tumour suppressor
XX cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
XX located on chromosome 13 at position 13q12-13q. AS3 has a role in
XX inhibiting cell proliferation and use as a marker for the efficient
XX diagnosis and treatment of prostate cancer. The invention includes AS3
XX cDNA and protein sequences, a vector comprising the cDNA sequence, a host
XX cell transfected with the expression vector, and a method for producing

CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC cytosolic activity, and acts to suppress cell proliferation. The AS3
CC gene is useful as a marker for the efficient diagnosis and treatment of
CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC antisense agents for sequence specific modulation of gene expression. The
CC AS3 protein may be used in the treatment of disorders caused by aberrant
CC modification or mutation of a gene encoding an AS3 protein, misregulation
CC of the AS3 gene or aberrant post-translational modification of the AS3
CC protein. This sequence represents the human AS3 cDNA sequence with an
CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC compared with the claimed AS3 cDNA sequence AAA28051.
XX
XX Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;
SQ

Query Match 100.0%; Score 5271; DB 21; Length 5355;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cggagagcccgagtgagcgagtagcgagtcggcaaccgaggggtagaatatatt 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
85 cggagagcccgagtgagcgagtagcgagtcggcaaccgaggggtagaatatatt 144
QY 61 ctgtcatggtctcattcaagactagagcaaatgatgaaaaattacatccgctggg 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
145 ctgtcatggtctcattcaagactagagcaaatgatgaaaaattacatccgctggg 204
QY 121 tcaaggaaatcacagataaaatctaaaggagagatggtgagacgattaaagatggtg 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
205 tcaaggaaatcacagataaaatctaaaggagagatggtgagacgattaaagatggtg 264
QY 181 tgaaaacttttatggatggaacagagactctgaagaaagaaaggactttattaaacc 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
265 tgaaaacttttatggatggaacagagactctgaagaaagaaaggactttattaaacc 324
QY 241 tagctttacatctgtctcagatttttttctcaagacatcctggaagatgttcgcttac 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
325 tagctttacatctgtctcagatttttttctcaagacatcctggaagatgttcgcttac 384
QY 301 tggtagcgtgctgctgctgatatatttcaggagatttatgctcctgaagctccttacacat 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
385 tggtagcgtgctgctgctgatatatttcaggagatttatgctcctgaagctccttacacat 444
QY 361 ccctgataaactaaaggatattttatgtttatacaagacagttgaaggggctagagg 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
445 ccctgataaactaaaggatattttatgtttatacaagacagttgaaggggctagagg 504
QY 421 atacaagagcccaaatcaatagatttttttatttacttgagaacattgcttgggtca 480
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
505 atacaagagcccaaatcaatagatttttttatttacttgagaacattgcttgggtca 564
QY 481 agtcataaacatagctttgagttagaagatagcaatgaaattttcaccagctataca 540
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
565 agtcataaacatagctttgagttagaagatagcaatgaaattttcaccagctataca 624
QY 541 gaaccttatttcagttataaacaatggccacaatcagaagtcacatgacatggtag 600
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
625 gaaccttatttcagttataaacaatggccacaatcagaagtcacatgacatggtag 684
QY 601 acctatgagctctattatttggaagtgatcacagtgctcagagcttttggatcagg 660
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
685 acctatgagctctattatttggaagtgatcacagtgctcagagcttttggatcagg 744
QY 661 ttttagtaaatctggtgacctgctcataagaattttaacaagcagcatatgattggcaa 720
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
745 ttttagtaaatctggtgacctgctcataagaattttaacaagcagcatatgattggcaa 804
QY 721 agcctttactgaagagagacgctcaagctattgagccatataattaccacttttttaac 780
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
805 agcctttactgaagagagacgctcaagctattgagccatataattaccacttttttaac 864
QY 781 aggttcgtgctgtggaaacatctcatcgatttgcagagcatgtctttgacttaa 840
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||

[illegible]

QY 2221 aaaaagaccocccctcaagccaaatgcccattcttgatctccatgcgatattttcta 2280
Db 2221 aaaaagaccoccccgctcaagccaaatgcccattcttgatctccatgcgatattttcta 2280
QY 2281 gtaaaagagaccagattgacagatattgagcctctgcataagagccctagatccaagca 2340
Db 2281 gtaaaagagaccagattgacagatattgagcctctgcataagagccctagatccaagca 2340
QY 2341 acctgaaacatctcaaacacattggttactattggttcataattgctctccttgacactg 2400
Db 2341 acctgaaacatctcaaacacattggttactattggttcataattgctctccttgacactg 2400
QY 2401 atcaatttgcctcttggaatacttggtagctacttttcaattgtagaagatcttctca 2460
Db 2401 atcaatttgcctcttggaatacttggtagctacttttcaattgtagaagatcttctca 2460
QY 2461 tgaatgatcgctctcagggaaaaagacaactaaaacttgggttccagatgaagaagtat 2520
Db 2461 tgaatgatcgctctcagggaaaaagacaactaaaacttgggttccagatgaagaagtat 2520
QY 2521 ctctgagacaattggtcaaaattcaggctattaaaatgatggttcgatggctacttggaa 2580
Db 2521 ctctgagacaattggtcaaaattcaggctattaaaatgatggttcgatggctacttggaa 2580
QY 2581 tgaataataatcacagtaaatcaggaacttctaccttaagattgctcaacaacaatttgc 2640
Db 2581 tgaataataatcacagtaaatcaggaacttctaccttaagattgctcaacaacaatttgc 2640
QY 2641 atagtatgagaaactgacagacaggggaaaaattagtaaacacagatatgtcacgtctga 2700
Db 2641 atagtatgagaaactgacagacaggggaaaaattagtaaacacagatatgtcacgtctga 2700
QY 2701 gacttgcctgctggagtgctatttgaagctggcacaagaacccctgttaccatgaaatca 2760
Db 2701 gacttgcctgctggagtgctatttgaagctggcacaagaacccctgttaccatgaaatca 2760
QY 2761 tcaattagaacaataatcagctatgtgcttagctatcaacgatcaagatgaatgctcaagtaa 2820
Db 2761 tcaattagaacaataatcagctatgtgcttagctatcaacgatgaatgctcaagtaa 2820
QY 2821 gacaagtgttgcacgaacttcaaaagcccttcccggttccagcttccacttgagt 2880
Db 2821 gacaagtgttgcacgaacttcaaaagcccttcccggttccagcttccacttgagt 2880
QY 2881 atatggcaatctgtgccccttctgcaaaagatcctgttaaaggagagaagctcatgcta 2940
Db 2881 atatggcaatctgtgccccttctgcaaaagatcctgttaaaggagagaagctcatgcta 2940
QY 2941 ggaatgtttgtgaaaaataataatgtaaggcggaggtatctggaagcagcatgcaagctg 3000
Db 2941 ggaatgtttgtgaaaaataataatgtaaggcggaggtatctggaagcagcatgcaagctg 3000
QY 3001 ttagtgaataattatgtctctcaccagatgatgtgttccatatacaaatccacttt 3060
Db 3001 ttagtgaataattatgtctctcaccagatgatgtgttccatatacaaatccacttt 3060
QY 3061 tggcacatgaccagattatgtcaaaagtacaggatattgaacaacttaaagatgttaaag 3120
Db 3061 tggcacatgaccagattatgtcaaaagtacaggatattgaacaacttaaagatgttaaag 3120
QY 3121 aatgtctttgttcttggaataatataatggtcgttaaaatgaaataacagtcacgctt 3180
Db 3121 aatgtctttgttcttggaataatataatggtcgttaaaatgaaataacagtcacgctt 3180
QY 3181 ttaacagaagatggttagaaattataacaacaagaagatgccacagcagatgatg 3240
Db 3181 ttaacagaagatggttagaaattataacaacaagaagatgccacagcagatgatg 3240
QY 3241 caaaaatgaatgaaaaactgtacactgtgtgtgattgttccatgaaatacatcatgtcaa 3300
Db 3241 caaaaatgaatgaaaaactgtacactgtgtgtgattgttccatgaaatacatcatgtcaa 3300
QY 3301 agagtactacatacagtttggaaatctcttaaaagaccoccggtactaccagctgttttctca 3360

Db 3301 agagtactacatacagtttggaaatctcttaaaagaccocgggtactaccagctcgtttctca 3360
QY 3361 ctcaacctgaagaatttcagtaaacacaaaaattatctgctcctcctgaaatgaaatcat 3420
Db 3361 ctcaacctgaagaatttcagtaaacacaaaaattatctgctcctcctgaaatgaaatcat 3420
QY 3421 ttttctactcttgaaaaacccataaaccaatgttcttagagctgtttaacaagccacttt 3480
Db 3421 ttttctactcttgaaaaacccataaaccaatgttcttagagctgtttaacaagccacttt 3480
QY 3481 catcagcaggaagaatctcagacccaaatcatcaagaattggaactgttaagaagaatgcaa 3540
Db 3481 catcagcaggaagaatctcagacccaaatcatcaagaattggaactgttaagaagaatgcaa 3540
QY 3541 gcaagcagctcaaatccaaagctctccttggaagaataaaggggaggtctgtagttctgaaa 3600
Db 3541 gcaagcagctcaaatccaaagctctccttggaagaataaaggggaggtctgtagttctgaaa 3600
QY 3601 tggatcacagtgaaaaatgaagattcacaaatgtcttcaccttgcgggggaaaaaagt 3660
Db 3601 tggatcacagtgaaaaatgaagattcacaaatgtcttcaccttgcgggggaaaaaagt 3660
QY 3661 acaagagagacgactctgttaaggtctgaattggaagagccttagagggcaggaaaa 3720
Db 3661 acaagagagacgactctgttaaggtctgaattggaagagccttagagggcaggaaaa 3720
QY 3721 aaagccctgcacagacagagaggaataataggatggtgacttgaactgaagtgtgtac 3780
Db 3721 aaagccctgcacagacagagaggaataataggatggtgacttgaactgaagtgtgtac 3780
QY 3781 aggaacagaaacctaaaggcagtcagcgaagtcggaaaagagccatcacggcttcagaat 3840
Db 3781 aggaacagaaacctaaaggcagtcagcgaagtcggaaaagagccatcacggcttcagaat 3840
QY 3841 ctgataaacagcagtgccctgagaaaaagaggtccaagaagataattagaaaaatgaag 3900
Db 3841 ctgataaacagcagtgccctgagaaaaagaggtccaagaagataattagaaaaatgaag 3900
QY 3901 atgaacagaaatagtcgcccaaaaaaggttaaagagagccaccaccaaacctcttggt 3960
Db 3901 atgaacagaaatagtcgcccaaaaaaggttaaagagagccaccaccaaacctcttggt 3960
QY 3961 gaggtacacccaaaagaagcgaacaaatgaaactcttcaaaaaggaagcaaaaaaat 4020
Db 3961 gaggtacacccaaaagaagcgaacaaatgaaactcttcaaaaaggaagcaaaaaaat 4020
QY 4021 ctggacctccagcacagagagaggaagaagaagaacaaagtggaaatcacgaac 4080
Db 4021 ctggacctccagcacagagagaggaagaagaagaacaaagtggaaatcacgaac 4080
QY 4081 agaagtccaaaaagcaaacagcagctgccaagggagacacagcagagcagaaatcttc 4140
Db 4081 agaagtccaaaaagcaaacagcagcagctgccaagggagacacagcagagcagaaatcttc 4140
QY 4141 ctgaaatctagtgcatttgaattccacacagtcacacacacagaaagacaggaagacct 4200
Db 4141 ctgaaatctagtgcatttgaattccacacagtcacacacacagaaagacaggaagacct 4200
QY 4201 caaaaagcgaatcacatcacacccaaaaaaatgtgtaagtgttaaatattacatttc 4260
Db 4201 caaaaagcgaatcacatcacacccaaaaaaatgtgtaagtgttaaatattacatttc 4260
QY 4261 aaaccaatttcaaatatttgcacaaagtctcctaaattgttaaacatacatattgtctga 4320
Db 4261 aaaccaatttcaaatatttgcacaaagtctcctaaattgttaaacatacatattgtctga 4320
QY 4321 ttttaattccatatatttagcccccattacactaggtcacggcggaagtgtcaaaagggga 4380
Db 4321 ttttaattccatatatttagcccccattacactaggtcacggcggaagtgtcaaaagggga 4380
QY 4381 agggcgatgaacaaaatgtaataaacttctctgaaagcttctggaagcttctggaataatctttt 4440

Db 1 ccgagagcccgaggtgagcggagtagcggagtcggccacccggagggttagaataattt 60
Qy 61 ctgtcattggtcattcaaaactaggaccatgatgaaataattacatatccgctggg 120
Db 61 ctgtcattggtcattcaaaactaggaccatgatgaaataattacatatccgctggg 120
Qy 121 tcaaggaaatcacagataaaatattcaaaaggagagatggtagacgattaaagatgtg 180
Db 121 tcaaggaaatcacagataaaatattcaaaaggagagatggtagacgattaaagatgtg 180
Qy 181 tgaataattttatggatatggaccagactctgaagaagaaaggagctttatttaaac 240
Db 181 tgaataattttatggatatggaccagactctgaagaagaaaggagctttatttaaac 240
Qy 241 tagctttacatctgtctcagattttttctcaagcactcgtgaaagatgttcgcttac 300
Db 241 tagctttacatctgtctcagattttttctcaagcactcgtgaaagatgttcgcttac 300
Qy 301 tggtagcctgtcgtcttgatatttttcaggattttatgctcctgaagcctcttacacat 360
Db 301 tggtagcctgtcgtcttgatatttttcaggattttatgctcctgaagcctcttacacat 360
Qy 361 cccctgataaactaaaggatatattttatgcttatacaagacagctgaaagggcagag 420
Db 361 cccctgataaactaaaggatatattttatgcttatacaagacagctgaaagggcagag 420
Qy 421 atacaagagcccaactcaatagatttttttatttacttgagaacattgcttgggtca 480
Db 421 atacaagagcccaactcaatagatttttttatttacttgagaacattgcttgggtca 480
Qy 481 agtcatataacatatgctttgagtagaagatgcaatgaaattttcacccagctataca 540
Db 481 agtcatataacatatgctttgagtagaagatgcaatgaaattttcacccagctataca 540
Qy 541 gaaccttatttccagttataaaacatggccacatcgaagtcocatgacatggttag 600
Db 541 gaaccttatttccagttataaaacatggccacatcgaagtcocatgacatggttag 600
Qy 601 acctatgagctctattttgtgaaggtgatacagtgctcagagcttttggatcgg 660
Db 601 acctatgagctctattttgtgaaggtgatacagtgctcagagcttttggatcgg 660
Qy 661 ttttagtaaatctggtaccctgctcataagaattttaaacaagcagcatatgattggcaa 720
Db 661 ttttagtaaatctggtaccctgctcataagaattttaaacaagcagcatatgattggcaa 720
Qy 721 aggccttactgaagagacagctcaagctattgagccatatattaccaccttttttaac 780
Db 721 aggccttactgaagagacagctcaagctattgagccatatattaccaccttttttaac 780
Qy 781 aggttctgagcttgggaaacatctacagcatttgcagacatgctttgaacttaa 840
Db 781 aggttctgagcttgggaaacatctacagcatttgcagacatgctttgaacttaa 840
Qy 841 ttttggagctctacaattatgatagtcatttgcctcctctgttttccccagcttgaat 900
Db 841 ttttggagctctacaattatgatagtcatttgcctcctctgttttccccagcttgaat 900
Qy 901 ttaaattaagagcaatgataatgagagcgccctacaagttgttaaaactctggcaaaa 960
Db 901 ttaaattaagagcaatgataatgagagcgccctacaagttgttaaaactctggcaaaa 960
Qy 961 tgtttgggcaaggattcagaaattgcttctcaaaaacagcactttgagctgact 1020
Db 961 tgtttgggcaaggattcagaaattgcttctcaaaaacagcactttgagctgact 1020
Qy 1021 tgggaggttttaatgatccatgataccaatccgcctgggaatgttgaaaattttgctagcc 1080
Db 1021 tgggaggttttaatgatccatgataccaatccgcctgggaatgttgaaaattttgctagcc 1080
Qy 1081 attgtctcatgaaccatcctgatttagcaaaagacttaacagagctatcttaaaagtgagg 1140
Db 1081 attgtctcatgaaccatcctgatttagcaaaagacttaacagagctatcttaaaagtgagg 1140

Qy 1141 cacatgacctgaggaagctatttagacatgatgttatttgtcaatagtacagctgcta 1200
Db 1141 cacatgacctgaggaagctatttagacatgatgttatttgtcaatagtacagctgcta 1200
Qy 1201 aaaaggatattcttctgttcaatgatcacttacttaatttttgtgagagagaaacattag 1260
Db 1201 aaaaggatattcttctgttcaatgatcacttacttaatttttgtgagagagaaacattag 1260
Qy 1261 acaaacgatggagagtacgcaaaagaagccatgatgggacttgcacaaatttataagaat 1320
Db 1261 acaaacgatggagagtacgcaaaagaagccatgatgggacttgcacaaatttataagaat 1320
Qy 1321 atgctttcacagtcagcgtggaaaagatgctcaaaacagatagcatggatcaaaagaca 1380
Db 1321 atgctttcacagtcagcgtggaaaagatgctcaaaacagatagcatggatcaaaagaca 1380
Qy 1381 aattgctacatatatatattcaaaaatagatttgatgactcgactacttgttgaacggatct 1440
Db 1381 aattgctacatatatatattcaaaaatagatttgatgactcgactacttgttgaacggatct 1440
Qy 1441 ttgctcaatcacatgggttctctcaaaatttagaaactacagaacggatgaaatgcttatatt 1500
Db 1441 ttgctcaatcacatgggttctctcaaaatttagaaactacagaacggatgaaatgcttatatt 1500
Qy 1501 acttgatgcccactggatttaaatgctgtgaagcatgaaatggaatgtggaaatgtc 1560
Db 1501 acttgatgcccactggatttaaatgctgtgaagcatgaaatggaatgtggaaatgtc 1560
Qy 1561 aaaaatctgctccgacatcaagtaagatttgcctgacttgatttaagcaaccccaaaacag 1620
Db 1561 aaaaatctgctccgacatcaagtaagatttgcctgacttgatttaagcaaccccaaaacag 1620
Qy 1621 atgcccagtgccaagggccatatatttcaaaagtgatggttattacaagaataattacctgact 1680
Db 1621 atgcccagtgccaagggccatatatttcaaaagtgatggttattacaagaataattacctgact 1680
Qy 1681 ctggttaaggtccaggtattcatgaagaaattccacacaggtgttagagatgatgagaaaa 1740
Db 1681 ctggttaaggtccaggtattcatgaagaaattccacacaggtgttagagatgatgagaaaa 1740
Qy 1741 taagaagagcagttagaagtagctttagtccaaactgctcctgcaagcagctgaaggtt 1800
Db 1741 taagaagagcagttagaagtagctttagtccaaactgctcctgcaagcagctgaaggtt 1800
Qy 1801 gtgtgcgtgaaataaactaaagaagtgggcaaccccaacacagcctacaaaatccttctctgg 1860
Db 1801 gtgtgcgtgaaataaactaaagaagtgggcaaccccaacacagcctacaaaatccttctctgg 1860
Qy 1861 aatgatacaagtttctctctgagagagtagcacctgtgcacatagatcacccaatctatca 1920
Db 1861 aatgatacaagtttctctctgagagagtagcacctgtgcacatagatcacccaatctatca 1920
Qy 1921 gtgctcttattaaacaagtgaaacaaatcaatagatgaaacagcagatgatgaagatgagg 1980
Db 1921 gtgctcttattaaacaagtgaaacaaatcaatagatgaaacagcagatgatgaagatgagg 1980
Qy 1981 gtgttcaactgatcaagccatcagagcaggtcttgaaactgttaaggtactctcattta 2040
Db 1981 gtgttcaactgatcaagccatcagagcaggtcttgaaactgttaaggtactctcattta 2040
Qy 2041 cacatccatctcatttctctgtgaaacatttgaatcatttactggtctgtctgaaaaa 2100
Db 2041 cacatccatctcatttctctgtgaaacatttgaatcatttactggtctgtctgaaaaa 2100
Qy 2101 tggatgataaaaaagtagcagaagctgactacaaaattttcaaaaacacagaggaagcaaaa 2160
Db 2101 tggatgataaaaaagtagcagaagctgactacaaaattttcaaaaacacagaggaagcaaaa 2160
Qy 2161 ttgaagagggattttccacacatcagatcagccttgccttctctgttttaacatcacaaaatcta 2220
Db 2161 ttgaagagggattttccacacatcagatcagccttgccttctctgttttaacatcacaaaatcta 2220

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:28:02 ; Search time 520.53 Seconds
(without alignments)
6358.268 Million cell updates/sec

Title: US-09-512-581-1

Perfect score: 5271

Sequence: 1 ccggagagcccgaggtag.....gtaaaaaaaaaaaaaa 5271

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5271	100.0	5271	21	AAA28051
2	5271	100.0	5355	21	AAA28052
3	1112.8	21.1	3957	21	AAA47423
4	411.4	7.8	439	20	AAV87629
5	409.4	7.8	738	20	AAZ15259
6	364	6.9	405	16	AAI19926
7	333	6.3	351	21	AA242861
8	292	5.3	295	21	AAO2817
9	258.8	4.9	284	20	AAZ14459
10	253.8	4.8	1217	21	AAF18338
c 11	203.6	3.9	530	21	AA280598

12	165.6	3.1	936	22	AAF58252	Oligonucleotide D1
13	165.6	3.1	936	22	AAF58254	Oligonucleotide D1
14	165.6	3.1	936	22	AAF58257	Oligonucleotide D1
15	165.6	3.1	936	22	AAF58259	Oligonucleotide D2
16	165.6	3.1	936	22	AAF58262	Oligonucleotide D2
17	165.6	3.1	938	22	AAF58255	Oligonucleotide D1
c 18	163.4	3.1	936	22	AAF58252	Oligonucleotide D1
c 19	163.4	3.1	936	22	AAF58254	Oligonucleotide D1
c 20	163.4	3.1	936	22	AAF58257	Oligonucleotide D1
c 21	163.4	3.1	936	22	AAF58259	Oligonucleotide D2
c 22	163.4	3.1	936	22	AAF58262	Oligonucleotide D2
c 23	163.4	3.1	938	22	AAF58255	Oligonucleotide D1
c 24	149.6	2.8	150	19	AAI12049	Human biallelic po
c 25	149.6	2.8	150	19	AAI12050	Human biallelic po
c 26	122	2.3	161	21	AAO6611	Human secreted pro
c 27	69.4	1.3	244	22	AAF58238	Oligonucleotide D1
c 28	67.2	1.3	244	22	AAF58238	Oligonucleotide D1
c 29	62	1.2	6644	20	AAI33181	Base sequence of t
c 30	62	1.2	7372	20	AAI33182	Base sequence of t
c 31	62	1.2	7797	20	AAI33180	Cowpox virus bar f
c 32	62	1.2	7996	20	AAI33184	Base sequence of t
c 33	58	1.1	6755	19	AAV21511	Staphylococcal bac
c 34	57.8	1.1	4000	18	AAI91902	Mannose-1-phosphat
c 35	55.2	1.0	5059	20	AAI84332	Stealth virus nucl
c 36	51.6	1.0	3399	17	AAI05868	Chicken leucocytos
c 37	51.4	1.0	72604	20	AAI10752	Genomic sequence o
c 38	50.8	1.0	1686	16	AAI087587	DNA encoding Leuco
c 39	49.2	0.9	1935	21	AAI99088	Human pancreatic c
c 40	48.6	0.9	396	22	AAI94862	Human ovarian canc
c 41	48.2	0.9	1236	21	AAI02163	Human colon cancer
c 42	48.2	0.9	10732	21	AAI10594	Gene encoding a su
c 43	48.2	0.9	49999	20	AAI23895	Murine LOBO homolo
c 44	47.6	0.9	6741	21	AAI10595	Gene encoding a su
c 45	47	0.9	6644	20	AAI33181	Base sequence of t

ALIGNMENTS

RESULT 1

AAA28051

ID AAA28051 standard; cDNA; 5271 BP.

XX

AC AAA28051;

XX

DT 01-DEC-2000 (first entry)

XX Human androgen shutoff gene 3 (AS3) cDNA sequence.

DE Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;

XX chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;

KW diagnosis; treatment; cytostatic; human; ss.

XX Homo sapiens.

OS

XX Key

FT exon

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

THIS PAGE BLANK (USPTO)


```

RESULT 14
AI911784/c 535 bp mRNA EST 17-DEC-1999
LOCUS wd32h01.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2329873 3', mRNA sequence.
ACCESSION AI911784
VERSION AI911784
KEYWORDS AI911784.1 GI:5631639
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 535)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 708 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 446.
FEATURES
Location/Qualifiers
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2329873"
/lab_host="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP.GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 205 a 90 c 81 g 159 t
ORIGIN
Query Match 10.0%; Score 528.6; DB 103; Length 535;
Best Local Similarity 99.3%; Pred. No. 5e-103; 4; Indels 0; Gaps 0;
Matches 531; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4723 actgacagtcaccgactgtttattggtactattgattgaaagaatttggtagataga 4782
|||||
Db 535 ACTGACAGTACCGAGCTGTTTATTGGACTATTGATTGAAAAGAATTGTTAGGATAGA 476
QY 4783 tcttaagcagtaactgtcagtggttattgtatttctgcaatttctactgtgaaaaa 4842
|||||
Db 475 TCTTAAGCAGTAATCTGTCAAGTGTGTTGTTATGTTCTGCAATTTACTGTGAAAAA 416
QY 4843 aaattgttttcaacaattgggtcatttcttgatgtcactatttggagaggttaa 4902
|||||
Db 415 AAATTTGTTTTTCAACAATTTGGTGTCATTTCTTGATGTCACATATTGTTGGAGAGTTAA 356
QY 4903 tggctctctcccttctgtgtatcttactagtggttactcctgggcacccttaattctcag 4962
|||||
Db 355 TGGTCTCTCTCCCTTGTGTATCTTACTAGTGTGTTTACTCCTGGCACCTTAACTCTCAG 296
QY 4963 aggtgctaataatgtctgcattacaccagaagatgcctctgatagaggaacaaccatgc 5022
|||||
Db 295 AGGTGCTAAATTTGTCGTCCATTACACAGAGGATGCCTCTGATAGGAGGACCAACCATGC 236
QY 5023 aaattgtgaatagtcctgaagttcttgattactttacacctcagattgttctcc 5082
|||||

```

```

Db 235 AAATTGTGAATAAGTCTCTGAAGTCTTGGATTACTTTACACCTCAGTATTGATTGTCTCC 176
QY 5083 agaatcttctggccttcatggcaatgaaattttaagaagaagatttaaaagtatttta 5142
|||||
Db 175 AGAATTTCTGGCCTTTCATGGCAATGAAATTTTAAAGAAGAAGATTTPAAGTATTTTA 116
QY 5143 attttaagaagtggtgtataaaaaataatgtactgaattcttattcccatcttcatcctt 5202
|||||
Db 115 ATTTTAAAGAGTGTGTATAAAATAATGACTGAATTTCTTTATCCCATTTTATCATCCTT 56
QY 5203 tcaagtttttataatctactgtatcatcaataaaaaattctgtaattggaatgaataaa 5257
|||||
Db 55 TCAGTTTTTATTAACTACTGTATCAATAAATCTGTAATTTTAAAGAGAAAAA 1
RESULT 15
BF111072 525 bp mRNA EST 20-OCT-2000
LOCUS 7n43b02.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3567291 3',
DEFINITION mRNA sequence.
ACCESSION BF111072
VERSION BF111072
KEYWORDS BF111072.1 GI:10940762
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 525)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 470.
FEATURES
Location/Qualifiers
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3567291"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 133 a 81 c 109 g 201 t
ORIGIN
Query Match 9.9%; Score 520.8; DB 144; Length 525;
Best Local Similarity 99.4%; Pred. No. 2.3e-101;
Matches 522; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4453 ttttggctcaagcttgaggctgaataaaagccttcttgatgcacaaaatgggactgtgaaga 4512
|||||
Db 1 TTTTGGTCAAGCTGAGGCTGAATAAAGCCCTTGTATGCACAAAATGGGACTGCTGAAGA 60

```



```
Best Local Similarity 99.3%; Pred. No. 1.6e-104;
Matches 538; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4715 ggcattaaactgacagcagcagctgtttattgagctattgattgaaagaatttgtt 4774
Db 542 GGCCTTAAATGACAGTACCGAGCTGTTTATGATGATTTGAAAGAAATTTGTT 483
QY 4775 agatagatcttaagcagtaactctgctgagttgttattgtatttctgcaatttact 4834
Db 482 AGGATAGATCTTAAGCAGTAATCTGTCAGTGTTTGTATTTGATCTCTCAATTTACT 423
QY 4835 gtgaaaaaaattgttttcaacaattggtgtcatttcttctgagtcactatttgggga 4894
Db 422 GTGAAAAAAATTTGTTTCAACAATTTGGTGCATTTTCTTGATGTCACATTTTGTGGA 363
QY 4895 gagttaaattggtctctccctctgtgtgtatcttactagtggttactctgggaccctta 4954
Db 362 GAGTTAAATGCTCTCTCCCTTTGTGTATCTTACTAGTGTACTCTGGGACCCCTTA 303
QY 4955 atctcagagtgctaaattgtctgccattacaccagagagatgctctgataggagac 5014
Db 302 ATCTTCAGAGTGCTAAATTTGTCGCCATTACACCAGAGGATGCCCTCTGATAGGAGAC 243
QY 5015 aaccatgcaaatgtgaaatagctgctgaagctcttggttacttttacacctcagattga 5074
Db 242 AACCATGCAAAATGTGAATAGTCTCAAGTCTCTTGGATTACTTTACACCTCAGTATTGA 183
QY 5075 ttgtcccaagaattttctgccccttcagcaatgaaataatttaagaagaagattttaa 5134
Db 182 TTTGTCCAGAAATTTCTGCCCCTTTCATGGCAATGAAATTTTAAAGAAGAAATTTAAA 123
QY 5135 gtatttcaattttaaagagtggttataaaatgactgaattcttccatttta 5194
Db 122 GTATTTTAAATTTAAAGAGTGTTTATAAATAATGATGTAATCTTTATCCCATTTTA 63
QY 5195 tcattccttcagtttttatttaatactactgtatcaataaaaaattctgtaatttgaatgagta 5254
Db 62 TCATCCTTTCAGTTTTTATTATTCCTCGTATCAATAAAATTCGTAAATTTGAATGAGAA 3
QY 5255 aa 5256
Db 2 AA 1

RESULT 13
BF142489 944 bp mRNA EST 24-OCT-2000
LOCUS 601788573F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016155 5',
DEFINITION mRNA sequence.
ACCESSION BF142489
VERSION BF142489.1 GI:10981529
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 944)
NTL-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: qgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM9263 row: a column: 20
High quality sequence stop: 620.
FEATURES
source 1..944
```

```
/organism="Mus musculus"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:4016155"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site: 1. NotI;
Site_2: SalI; transgenic model WNT-1, expression driven by
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dr. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 277 a 221 c 218 g 228 t
ORIGIN
Query Match 10.1%; Score 532.8; DB 145; Length 944;
Best Local Similarity 86.7%; Pred. No. 6.8e-104;
Matches 633; Conservative 0; Mismatches 92; Indels 5; Gaps 4;
QY 2557 tgatggttcagtgctacttggaatgaaaaataatcacagtaaatcacgaacttctacct 2616
Db 1 TGAATGTTCCATGCTACTTTGGAATGAAGATAATACAGTAAGTCAGGAACTTCCACCC 60
QY 2617 taagattgctaaacaacaattatgcatagtgatggagacttgacagacaggggaaaaatta 2676
Db 61 TCAGATTACTTAACACAGATACTGCTAGTAGTATGGGATTTGACAGAACAGGGGAAATTA 120
QY 2677 gtaaacagatatgtcacgtctgagactgtctgtggagtgctattgtgaagctggcac 2736
Db 121 GTAACCCAGATATGTCACGCTGAGACTTGTCTGGAGTGCCTATTGTGAAGCTGGCAC 180
QY 2737 aagaacctgttaccatgaaataacacattagaaataatcacactatgctcattagcta 2796
Db 181 AGGAGCCCTGTTACCCAGAGATCAATACACTGGAGCATACCAGCTGTGTGCATAGGCCA 240
QY 2797 tcaacgatgaatgctatcaagtaagacaagtggttggccagaatactcacaaagccctt 2856
Db 241 TCATGATGAGTGTATCAAGTCAGCAGGAGTGTGCTGCAGAAATTCACAAAGCCCTTT 300
QY 2857 cccgtttacggcttccacttgagtatatggcaatctgtgccctttgtgcaaaagactgtg 2916
Db 301 CCCGCTTACGGCTTCCCTTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAGACCCCTG 360
QY 2917 taaaggagagaagagctcatgctagggcaatgtttgggtaaaaaataataatgtaaggcggg 2976
Db 361 TGAAGAGAGAGCGGAGCCCATGCTAGACAGTGTCTGGTGAAGAACATCACTGTGAGGAGGG 420
QY 2977 agtatctgaagcagcagctgctgttagtgaataattattgtctcttctaccagagtatg 3036
Db 421 AGTACCTGAAGCAGCATGTCAGCTGTTAGTGAATAATTTATTGTCTCTTCTACCAAGTATG 480
QY 3037 ttgttccatatacaattcaacttttggcacatgaccacagattatgtcaaaagatcacagata 3096
Db 481 TGGTTCATATACAAATTCACCTTTGGGCACATGACCCAGATTTATGTCAACGTACAGGATA 540
QY 3097 ttgaacaacttaagaagttaaaagaatgtcttggtttcttctggaataataatagcta 3156
Db 541 TTGAACAACCTTAAAGATGTGACAAGAATGCTCTGG-TTGTCTCTGGAGATATTGATGGCTA 599
QY 3157 aaaaataaacaagtcacgcttttcatcagaagaatggtagaataattataaacaacaa 3216
Db 600 AAAATGAACAAACAG-CATGCTTTTATCAGAAAATGGTAGAAAATATTAAACAGACCAA 658
QY 3217 aagatgccccagaccagatgatacaaaaaatgaatgaaaaaactgtacactgtgtgatg 3276
Db 659 AGATTGCTCCAGGACCGATGAT-CACCACCTGACTGCACAACTGT--CCCCGTGTGTGATG 715
QY 3277 ttgccaatgaa 3286
Db 716 TTGCCATGGA 725
```



```

|||||
Db 481 CAGATGATGAGAAATGAATGAAAACTGTACACTGTGTGTGATGNGCCATGAATATCA 540
QY 3292 tcatgtcaagagtactacatac-agtgtggaatctcc 3328
|||||
Db 541 TCATGTCAAAGAGTACTACATACNAGTTGGATCTCC 578
|||||

RESULT 11
BF002153/c
LOCUS 7h23a12.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:3316798 3',
DEFINITION mRNA sequence.
ACCESSION BF002153
VERSION BF002153.1 GI:10702428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 463.
FEATURES
source
location/Qualifiers
1..540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3316798"
/clone_lib="NCI_CGAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 209 a 93 c 85 g 153 t
ORIGIN

Query Match 10.2%; Score 536.8; DB 143; Length 540;
Best Local Similarity 99.6%; Pred. No. 8.6e-105;
Matches 538; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4711 ctgcgcttaactgacagtagtaccgactgtttattggtatctattgattgaaagaatt 4770
|||||
Db 540 CTGCGCGCTTTAAACTGACAGTACCCGACGTGTTTATTGGATCTATTGATTTGAAAGAATT 481
CTGCGCGCTTTAAACTGACAGTACCCGACGTGTTTATTGGATCTATTGATTTGAAAGAATT 481
QY 4771 tattaagatagatcttaagcagtaactctcagtggtttattgatttcttctgcaattt 4830
|||||
Db 480 TGTAGTAGTAGATCTTAAAGCAGTAATCTGTACGTGTTGTTGTTGTTCTCTGCAATTT 421
TGTAGTAGTAGATCTTAAAGCAGTAATCTGTACGTGTTGTTGTTGTTCTCTGCAATTT 421
QY 4831 tactgtgaaaaaaattgttttccaacaattggtgtcattttcttctgattgtcactattgt 4890
|||||
Db 420 TACTGTGAAAAAAATTTGTTTCACAAATGCTGTGCATTTCTTGATGTCACATATTGT 561
TACTGTGAAAAAAATTTGTTTCACAAATGCTGTGCATTTCTTGATGTCACATATTGT 561

```

```

QY 4891 tggagagttaaatggtctctctccctttgtgtatcttacctagtggtttactcctgggaacc 4950
|||||
Db 360 TGGAGAGTTAAATGGTCTCTTCCCTTTGTGTATCTTACCTAGTGTCTTACTCTCTGGGCACC 301
|||||
QY 4951 cttaatcttcagaggtgctaaattgtctccattacaccagaagagtcctctgatagga 5010
|||||
Db 300 CTTAAATCTTCAGAGGTGCTAAATGTCTGCCATTACACCAGAAGGATGCTCTGTATAGGA 241
|||||
QY 5011 ggacaaccatgcaaatgtgaaatagtcctggaagtcttggaattctgtattactttacacctcagta 5070
|||||
Db 240 GGACACCATGCAAAATGTGAATAGTCTCTGAAGTCTTTGGATTACTTTACACCTCAGTA 181
|||||
QY 5071 ttgattgtccagaattttcttgccttttccatggaatgaaatttttaagaagaagatt 5130
|||||
Db 180 TTGATTGTGCCCAGAAATTTCTGCCTCTTCATGCCAATGAAAAATTTTAAAGAAGAAATT 121
|||||
QY 5131 taaagtatttttaatttaagagtggtgtatataaaataatgactgaattctttatcccat 5190
|||||
Db 120 TAAAGTATTTTAAATTTAAAGAGTGTGTATAAAATTAATGACTGAATCTTTATCCCAT 61
|||||
QY 5191 ttatcatcctttcagtttttatttaatactactgtatcaataaaattctgttaattgaaatg 5250
|||||
Db 60 TTTATCATCTTTCAGTCTTTTATTATTAATCCACTGATCAATAAAATCTGTAAATTGAATG 1

RESULT 12
LOCUS AW303630/c
DEFINITION XV21e03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2813788 3', mRNA sequence.
ACCESSION AW303630
VERSION AW303630.1 GI:6713319
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
FEATURES
source
location/Qualifiers
1..542
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2813788"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 208 a 93 c 84 g 157 t
ORIGIN

Query Match 10.2%; Score 535.6; DB 114; Length 542;

```



```

Db 301 TCCTCACAATTTAGAACTACAGACGGATGAATGCTTATATTACTGTGTATGCCACACT 360
QY 1517 ggatttaaatgctgtgaagcattgaatgaatgtgaaatgtgaaatgtcctccgaca 1576
Db 361 GGATTTAAATGCTGTGAAGCAATTTGAATGACATGTGGAATGTCAAAATCTGCTCCGACA 420
QY 1577 tcaagta-aaggattgctgtgaattgattgaagc-aaccccaaaacagatgcagtgctcaag 1634
Db 421 TCAAGTACAAAGGATTTGCTTGACTTGATTAAAGCAAAACCCAAACACAGATGCCAGTGTCAAG 480
QY 1635 gccatatttc-aaaagtgtgtattacaagaatttaccctgatacctgtgtaaggctca 1693
Db 481 GCCATATTTCAAAAGTGTGTGTATTACAAAGAAATTTACCTGTGATGCTGCTGTAAGGCTCA 540
QY 1694 ggatttcgaagaataacacacaggtgtgttagaaga-----tgatgagaataaagaaga 1747
Db 541 GGATTTTCATGAAGAAATTCACACAGGTGTTTCAGACAGATGACTGAGGACAAATCAAGAAA 600
QY 1748 gcaatttag-aagtaactgttagtccaaatgctcctgc--aagcaggctgaagggtgtgt 1804
Db 601 GCAGTTAGAAAGTACTGTGTGTAGTCCAAACATGCTCCTGCCAAAGCGAGGCTGACAGTGTGT 660
QY 1805 gcgtgaataactcaagaagtgtggcaccaccccaacacagctcacaaatccttctcctggaat 1864
Db 661 GCCTGACATAACCTAGAGTTGGGCAACCCCAACACAGATACAGATC-TTCCTGGCAAT 719
QY 1865 gatcaagtttctcttgagagagatgacac 1893
Db 720 GCTCAGCGTCTCTCTGGAGACGATAGTCTC 748

RESULT 8
AW976150 738 bp mRNA EST 02-JUN-2000
LOCUS EST388259 MAGE resequencences, MAGN Homo sapiens cDNA, mRNA sequence.
DEFINITION AW976150
ACCESSION AW976150
VERSION AW976150.1 GI:8167375
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 738)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 357
Seq primer: Forward.
Location/Qualifiers
1. .738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cclone_lib="MAGE resequencences, MAGN"
/note="Vector: pBluescriptSKm"

BASE COUNT 214 a 126 c 129 g 269 t
ORIGIN

Query Match 11.1%; Score 583.6; DB 122; Length 738;
Best Local Similarity 99.3%; Pred. No. 8.2e-115;
Matches 586; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4674 atgttctcattcgaatgctgtgatttagctttttagctgagcttaactgaacagctac 4733
|||||

```

```

Db 1 ATCTTTTGTGATTTCTGAAGTCTGTGTATAGCTTTTATCTCGCGCTTTAAACTGCACGTAC 60
QY 4734 ccgactgtttattggatctctattgatttgaagaatattgttagatagatcttaagcagt 4793
Db 61 CCACACTGTTTATTGGATCTATTGATTTGAAAGAGAAATTTGTTAGGATAGATCTTAAGCAGT 120
QY 4794 aatctgcagtggtttgtattgttatttctgcaatttactgtgaaaaaaatattgtttt 4853
Db 121 AATCTGTGACGTGTTGATTTGATTTCTCTGCAATTTTACTGTGAAAAAAATTTGTTT 180
QY 4854 caacaattggtgtcatttcttctgacatgtcactatttctgtggagagattaaatggtctcttc 4913
Db 181 CACAATTTGTTGTCATTTTCTTGATGTCACTATTGTTGGAGAGTTAAATGGTCTCTTCC 240
QY 4914 ctttgtatcttaccctagtggttactcctggcacccttaactcttcagaggtgtcctaaat 4973
Db 241 CTTTGTGATCTTACCTAGTGTTTACTCTCTGGCACCTTTAATCTTCAGAGGTGCTAAAT 300
QY 4974 tctctgccattacaccagaaggatgctctgtagaggagacacccatgcataattgtgaaa 5033
Db 301 TGCTGTCATTACACCACAAAGGATGCTCTGATAGGAGGACAAACATGCAAAATTGTGAAA 360
QY 5034 tagtctgaagttcttggtattcttaccctcagtcattgttcccgaaatttctg 5093
Db 361 TAGTCTGAAGTTCTTGGATTACTTTTACACCTCAGTATTGATTTGTCCTCCAGAAATTTCTG 420
QY 5094 gccttctcaggaatgaaatattttaaagaagaagattttaaagattttaaagag 5153
Db 421 GCCTTTTCATGCAATGAAATTTTAAAGAACAAAGATTTAAAGTATTTTAAAGAG 480
QY 5154 tgtgtataaataatgtactgaatttcttccatccatttccatccatttccatccatttccat 5213
Db 481 TGTGTTATAAAATAATGTACTGAATTTCTTTATCCCATTTTATCATCTCTTCAGATTTTAT 540
QY 5214 taatctactgtatcaataaattctgtaatttgaatgaagtaaaaaaaa 5263
Db 541 TAATCTACTGTATCAATAAAATTTCTGTAATTTGAATGAGAAAAA 590

RESULT 9
AW629426/c 568 bp mRNA EST 31-MAR-2000
LOCUS hl57b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2976383 3', mRNA sequence.
ACCESSION AW629426
VERSION AW629426.1 GI:7376216
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 568)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
Location/Qualifiers
1. .568
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cclone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site:1; Not I; Site:2; Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
source"

```


Db	541	TCATGTCAAGAGTACTACATACAGTTTGGAAATCTCTAAAGACCCGGTACTACCA	596
RESULT	7		
LOCUS	BG387646		
DEFINITION	802 bp	mRNA	EST
DEFINITION	602412496F1 NIH_MGC_92	Homo sapiens cDNA clone	IMAGE:45210
DEFINITION	mRNA sequence.		
ACCESSION	BG387646		
VERSION	BG387646.1	GI:13281092	

VERSION
KEYWORDS
SOURCE

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 802)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
AUTHORS
REFERENCE
TITLE
COMMENT
Contact: Robert Strausberg, Ph.D.

found through the I.M.A.G.E.E. CONSORTIUM
<http://image.llnl.gov>
 Plate: LLAM10419 row: p column: 24
 High quality sequence stop: 637.

FEATURES

```

/clone= IMAGE:4521035
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab host="NIH3T3 (phage-resistant)"

```

/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Lib

270 a	149 c	167 g	216 t
-------	-------	-------	-------

Query Ma
Best Loc
Matches

Qy 1158

QY 1218

Qy 1278

QY 1337

Qy 1397

QY 1457

QY 2792 agctataacagatgaatgtctatcaagttaagacagaagtgtttggccagagaacattcacaaagg 2851
 Db 480 AGCTATCAAGCATGAATGTCTATCAAGTAAGACAAAGTGTGGCCAGAAAACCTTCACAAAG 539
 QY 2852 cctttccatttac-ggcctccacttgagtatatgcaatctgtgcccctttgtgcaa-aa 2909
 Db 540 CTTTCCCGTTTACGGGCTTCCACTTGGATATATGCAATCTGTGCGCTTGGTGCAAGAA 599
 QY 2910 gatcctgtataa---ggagagaagagctcatgctgtaggcaatgtttgtggaataataaaat 2966
 Db 600 GATCTGCTTAAGAGAGAGAAAGAGCTCATGCTAGGCAATGTTTGGTGAACAATATAAAT 659
 QY 2967 --gtagggcgggagtagtctgaaagacagatgacgtcttagtgaataattattgtctcttc 3024
 Db 660 TGTACAGGGGGAGGATCTCTTGAAGCAGATTTGGGCTGTTAGTGAGAACTTATGGGCTTC 719
 QY 3025 taccagagtagtgtttccatatacaattacacctttgtgacacatgacc 3073
 Db 720 TACCACAG-ATTGGTGCCCTATATGATTCACCTTTTGGGCGCTTGGCCC 767

RESULT 5

BG432960 729 bp mRNA EST 14-MAR-2001
 LOCUS 602495783F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4609533 5',
 DEFINITION mRNA sequence.

ACCESSION BG432960
 VERSION BG432960.1 GI:13339466

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC

TITLE http://imgc.ncl.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-re@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: LNCM1353 row: e column: 22

High quality sequence stop: 722.

Location/Qualifiers

FEATURES

source

1..729

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4609533"

/clone_lib="NIH_MGC_75"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:

SfiI (ggcgcctcgcc); Site_2: SfiI (ggcattatggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGCCCATATGCGC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCCGCGCGCGCCGATG-dT(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.65

kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

BASE COUNT 225 a 129 c 144 g 231 t

ORIGIN

Query Match

Best Local Similarity 11.6%; Score 612.4; DB 153; Length 729;

Matches 715; Conservative 0; Mismatches 6; Indels 9; Gaps 9;

QY 428 gagccccacaattcaataggtatttttttacttgagacatgcttggtgcaagtcaata 487
 Db 1 GAGCCACACAGTTCATAGTATGTTTATTTACTTTGAGAACA-TGCTTGGGTCAAGTCATA 59
 QY 488 taacatatgctttgagttagaagatagcaatgaaattttcacccagctatcacagaacctt 547
 Db 60 TAACATATGCTTTTGGTGTAGAGTAGCAATGAAATTTTCCACCAGCTATACAGAACCTT 119
 QY 548 attttcagttataaacaatggccacacatcagaagtcacatatgcacatggtagaccttat 607
 Db 120 ATTTTCAGTTATTAACAATGGCCACCAATCAGAAAGTCCATATGCACATGCTAGACCTTAT 179
 QY 608 gagctctattatttgaagggtgatacagtgctcagggagcttttggatagcaggttttagt 667
 Db 180 GAGCTCTATTATTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTAGT 239
 QY 668 aaatctgggtacctgctcataagaatttaaacagaagcagcatatgatttggcaaggcttt 727
 Db 240 AAATCTGGTACCTGCTCATTAAGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGCTTT 299
 QY 728 actgaagagagacagctcaagctattgagccatatattaccacatttttttaacagggtctt 787
 Db 300 ACTGAAGAGAGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTAAATCAGGTTCT 359
 QY 788 gatgcttgggaaacatctatcagcgatattgtcagagcatgtc-tttagcttaatttgg 846
 Db 360 GATGCTTGGGAAAACATCTATCAGCGATTTGTGAGAGCATGTCTTTGACTTAATTTGG 419
 QY 847 agctctacaattatgatactcattgctctctctctctctctctctctctctctctctctct 905
 Db 420 AGCTCTACAAATATTGATGATGATTTGCTGCTCTCTGTTTACCCAGCTTTGAAATTTAA 479
 QY 906 ttaaaagagcaatgataatgaggagcgccctacaagtggttaaaactactggcaaaaatgttt 965
 Db 480 TTAAGAGCAATGATAATGAGGAGCGCTACAAGTTGTTAAACTACTGCGCAAAAATGCTTT 539
 QY 966 ggggcaaggagattcagaattggtcttc-aaacaagccacctttggcagtgctac-ttgg 1023
 Db 540 GGGGCAAGGATTCAGAAATTTGGCTTCTCAAAAACAAGCCACTTTGGCAGTGCTACTTTGG 599
 QY 1024 gcagggttaatgatataccatgataccaa-tccgcctggaatgtgga-aatttgcagacca 1081
 Db 600 GCAGGTTTAAATGATATCCATGATACCAATTCGCCCTGGAATGTGTGACCATTTCGTAGCCA 659
 QY 1082 ttgtctcagaaccatcctctgatttagc-aaagacttaacagagatctctt-aaagttagg 1139
 Db 660 TTGCTCATGAACCATCTTGATTTAGCAAAAAGATTTTAAACAGAGATCTCTTAAAGTCAGG 719
 QY 1140 tcaatgacc 1149
 Db 720 TCACCTTGAC 729

RESULT 6

BE504550 596 bp mRNA EST 04-AUG-2000
 LOCUS hz58a05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3212144 3'
 DEFINITION similar to TR:094237 094237 HYPOTHETICAL 45.2 KD PROTEIN ;, mRNA
 sequence.
 ACCESSION BE504550
 VERSION BE504550.1 GI:9706958
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 596)
 AUTHORS NCI-CGAP
 TITLE http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-re@mail.nih.gov

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA-yes.

Location/Qualifiers

source

1. 791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3086220"
/clone_lib="NCI-CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="vector: pT73b-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; NCI-CGAP_Sub8 is a subtracted library derived from NCI-CGAP_Sub5. The NCI-CGAP_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI-CGAP_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI-CGAP_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population); a pool of clones from NCI-CGAP_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population); NCI-CGAP_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI-CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550); 25% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Approaches (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_LIB=NCI-CGAP_Lu5
TAG_TISSUE=lung
TAG_SEQ=CAAC"
BASE COUNT 270 a 151 c 144 g 226 t
ORIGIN

Query Match 13.6%; Score 716; DB 149; Length 791;
Best Local Similarity 99.6%; Pred. No. 3, le-143;
Matches 749; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

Qy 2745 tttaccatgaatcacattagacaataatcagctatgtgcattagctatcaacgat 2804
Db 13 TGTACCATGAATCATCATATAGACAATATACAGTATGTGCATTAGCTATCAACGAT 72
Qy 2805 gaatgctatgaagaagaagttgtcccaagaacttcacaaaggctttccggttta 2864
Db 73 GAATGCTATCAAGTAAGCAAGTGTTCGCCAGAAACTTCACAAAGGCTTTCCGCTTTA 132
Qy 2865 cggctccacttgatgatgcaatctgtgcctttgtgcaaaagactctgtaagagag 2924
Db 133 CGGCTTCACATTCAGTATATGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAG 192
Qy 2925 agaagagctcatgctaggcaaatgtttgtgtaaaaaataataatgtaaggcggtgatcttg 2984
Db 193 AGAAGAGCTCATGCTAGGCAATGTTTGTGTAATAATAAATGTAAGCGGGAGTATCTG 252
Qy 2985 aagcagcagctgagctgttagtgaaaaaattattgtctctaccagagatattgtttcca 3044
Db 253 AAGCAGCATGAGCTGTTAGTGAATAAATATTGTCTCTCTACAGAGTATGTTGTTCCA 312
Qy 3045 tatacaattcacttttggcacatgacccagattatgtcaagtagcacagatattgaacaa 3104
Db 313 TATACAAATTCACCTTTTGGCAGATGACCCAGATATATGTCAAAGTACAGGATATTGAACAA 372
Qy 3105 cttaagaagtgtaagaagatgtctttgtttgtcttggaataattattggtgtaaaatgaa 3164
|||||

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	716	13.6	791	149	BF509252	UI-H-B14-BF509252
2	663.4	12.6	734	175	BG256731	602370939
3	632	12.0	648	122	AW973437	EST385430
4	621.2	11.8	848	141	BE973840	601483992
5	612.4	11.6	729	153	BG432960	602495783
6	596	11.3	596	136	BE504550	hz58a05.x
7	591.4	11.2	802	153	BG387646	602412496
8	583.6	11.1	738	122	AW976150	EST388259
9	560.4	10.6	568	118	AW629426	hi57b12.x
10	559.2	10.6	578	120	AW770571	hl86f11.x
11	536.8	10.2	540	143	BF002153	7h23a12.x
12	535.6	10.2	542	114	AW303630	xv21e03.x
13	532.8	10.1	944	145	BF142489	601788573
14	528.6	10.0	935	103	AI911784	wd32h01.x
15	520.8	9.9	525	144	BF111072	7n43b02.x
16	503.8	9.6	608	32	AV685306	AV685306
17	497.4	9.4	608	149	BF471938	UI-M-BH3-BF471938
18	495	9.4	683	136	BE532986	601235013
19	482.8	9.2	486	23	AI681422	tx46e01.x
20	482.6	9.2	535	137	BE551003	7b67b03.x
21	479.4	9.1	888	155	BG571128	602591748
22	478.4	9.1	480	23	AI655429	ts98f06.x
23	474.8	9.0	490	149	BF511856	UI-H-B14-BF511856
24	467.4	8.9	469	9	AA563884	nk19e08.s
25	466.8	8.9	470	4	AA282190	zs99b08.r
26	456.2	8.7	461	15	AI016896	ou31c08.x
27	437.2	8.3	442	7	AA447433	zw93f10.x
28	430.4	8.2	433	18	AI275866	q168c06.x
29	423	8.0	423	23	AI680124	tw65b01.x
30	407.4	7.7	409	9	AA625961	zu92e08.s
31	406.8	7.7	410	14	AA987361	or92g04.s
32	406.8	7.7	648	137	BE570606	601329907
33	402.8	7.6	406	18	AI299646	qnl2a10.x
34	393	7.5	405	16	AI140790	ocl5b04.x
35	392	7.4	434	116	AW483768	56376 MAR
36	389	7.4	429	145	BF189753	235446 MA
37	388.4	7.4	390	148	BF446410	7p36h09.x
38	388.4	7.4	391	10	AA630616	ac11e07.s
39	379.2	7.2	735	154	BG532020	602561049
40	377	7.2	445	142	BE948523	UI-M-BH3-BE948523
41	376.4	7.1	738	105	AL045878	DRFp434J
42	373.4	7.1	580	149	BF451878	uz80g04.y
43	373.4	7.1	738	32	AV710229	AV710229
44	368.2	7.0	595	143	BF043498	BP250017A
45	365	6.9	703	112	AW175091	fi32b08.Y

ALIGNMENTS

RESULT 1
LOCUS BF509252
DEFINITION UI-H-B14-aow-c-07-0-UI.s1 NCI-CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3086220 3', mRNA sequence.
ACCESSION BF509252
VERSION BF509252.1 GI:11592550
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 791)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

117: gb_est48: *
118: gb_est49: *
119: gb_est50: *
120: gb_est51: *
121: gb_est52: *
122: gb_est53: *
123: gb_est54: *
124: gb_est55: *
125: gb_est56: *
126: gb_est57: *
127: gb_est58: *
128: gb_est59: *
129: gb_est60: *
130: gb_est61: *
131: gb_est62: *
132: gb_est63: *
133: gb_est64: *
134: gb_est65: *
135: gb_est66: *
136: gb_est67: *
137: gb_est68: *
138: gb_est69: *
139: gb_est70: *
140: gb_est71: *
141: gb_est72: *
142: gb_est73: *
143: gb_est74: *
144: gb_est75: *
145: gb_est76: *
146: gb_est77: *
147: gb_est78: *
148: gb_est79: *
149: gb_est80: *
150: gb_est81: *
151: gb_est82: *
152: gb_est83: *
153: gb_est84: *
154: gb_est85: *
155: gb_est86: *
156: gb_est87: *
157: gb_est88: *
158: gb_est89: *
159: gb_est90: *
160: gb_est91: *
161: gb_est92: *
162: gb_est93: *
163: gb_est94: *
164: gb_est95: *
165: gb_est96: *
166: gb_est97: *
167: gb_est98: *
168: gb_est99: *
169: gb_est100: *
170: gb_est101: *
171: gb_est102: *
172: gb_est103: *
173: gb_est104: *
174: gb_est105: *
175: gb_est106: *
176: gb_est107: *
177: gb_est108: *
178: gb_est109: *
179: gb_est110: *
180: gb_est111: *
181: gb_est112: *
182: gb_est113: *
183: gb_est114: *
184: gb_est115: *
185: gb_est116: *
186: gb_est117: *
187: gb_est118: *
188: gb_est119: *
189: gb_est120: *

190: gb_est121: *
191: gb_est122: *
192: gb_est123: *
193: gb_est124: *
194: gb_est125: *
195: gb_est126: *
196: gb_est127: *
197: gb_est128: *
198: gb_est129: *
199: gb_est130: *
200: gb_est131: *
201: gb_est132: *
202: gb_est133: *
203: gb_est134: *
204: gb_est135: *
205: gb_est136: *
206: gb_est137: *
207: gb_est138: *
208: gb_est139: *
209: gb_est140: *
210: gb_est141: *
211: gb_est142: *
212: gb_est143: *
213: gb_est144: *
214: gb_est145: *
215: gb_est146: *
216: gb_est147: *
217: gb_est148: *
218: gb_est149: *
219: gb_est150: *
220: gb_est151: *
221: gb_est152: *
222: gb_est153: *
223: gb_est154: *
224: gb_est155: *
225: gb_est156: *
226: gb_est157: *
227: gb_est158: *
228: gb_est159: *
229: gb_est160: *
230: gb_est161: *
231: gb_est162: *
232: gb_est163: *
233: gb_est164: *
234: gb_est165: *
235: gb_est166: *
236: gb_est167: *
237: gb_est168: *
238: gb_est169: *
239: gb_est170: *
240: gb_est171: *
241: gb_est172: *
242: gb_est173: *
243: gb_est174: *
244: gb_est175: *
245: gb_est176: *
246: gb_est177: *
247: gb_est178: *
248: gb_est179: *
249: gb_est180: *
250: gb_est181: *
251: gb_est182: *
252: gb_est183: *
253: gb_est184: *
254: gb_est185: *
255: gb_est186: *
256: gb_est187: *
257: gb_est188: *
258: gb_est189: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 14:25:52 ; Search time 6640.39 Seconds
(without alignments)
7503.474 Million cell updates/sec

Title: US-09-512-581-1
Perfect score: 5271
Sequence: 1 ccgagagcccgagtgag.....gtacaaaaaaaaaaaaa 5271

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

Db 1451 erleikdrdyfelyvtantnlvegkletplqadheedsidrrseemeikvlgeklernqy 1510
Qy 260 -----LELYN-----IDSHLLSVL--POLEFKLKs----- 283
Db 1511 llerlqeekfelsnkleilqkemetsvllkddlqgklesliseniilkenidttlkhhssd 1570
Qy 284 -----NONEERLQVVKLLAKMFGAKDELASQNK-----PLWQCYVGRFNDIH----- 326
Db 1571 tqaglkgtqgelqlaknla--iaasdnpcitqeketsadcvhpleeekillilteelhgktn 1628
Qy 327 -----VPRLECVKASCLMHPDLADLTLEYLKVSRHDPEE--ATRHD 369
Db 1629 egekllehekneleaqvelkce--veh-lmksmieskssleslqnekhdtteqllalqkq 1685
Qy 370 VIVSIVTAAKKDIIIVNDHL-----LNF-----VRERTLOKRWVRKEAMMGL 412
Db 1686 --mqvvtqekkelqgthelhtaevdhikeniqlgnfkneaqgkttkeqcllnenkeleq 1743
Qy 413 AQIYKKYALQSAAGKADRAKQIAWIKDK-----LHIIYQNSIDDRLLVERIFAYVMVPHN 467
Db 1744 sq-----hrlqc-----eieelmkslkdesaletikeseqvinlnqenemvmlmeeikn 1795
Qy 468 LETT---ERMKCLYLYATLDLNAVKALNEMWKQNLRLHQVKDLDLLIKQ----- 515
Db 1796 sqrtviaeqldldresvem-sietqddlrkaeqalqgdkdvqeltsqlsvlqekis 1854
Qy 516 -----PKTDASVKAIFSKVMVITRNLPDPGKA-----QDFMK 547
Db 1855 llenqmylnvatvketiserddlqgkqhlfsietislsikefaleqaeakadaar 1914
Qy 548 KFTQVLEDEDEKIRKOL-----EVLVSPTCSCKQAGC-----VREITTK--- 586
Db 1915 ktiditekisnieeqllqgatknetiyyeresli-----qckeqlalntehliretlkskdl 1970
Qy 587 -LGPNKQPTNPFLEKIFELLERAPVHIDTESISALIKQVKNKSIDGTADDEGVPTDQA 645
Db 1971 alghmeqerdeaanvial-----tekmslsleeqinenvttikegege----- 2033
Qy 646 IRAGLELLKVLSTFPHISFSAETTESLACLKMDDEKVAEALQIFKNTG----- 696
Db 2014 -----ketfylrpskqssqmeelreslktkdqlqleaakeiseatneiknlaki 2066
Qy 697 SKIEDPPIHRSAL-LPVLHHSKKGPPRQAKYAIHCHAIFFSSKETQFAIFEPHLKSL 755
Db 2067 slleeeilqnasilneavserenlrhskqqlvseleqlsltlksrdhafaq-----skr 2120
Qy 756 DPSNLEHLITPLVTIGHTALLAPDQFAAPKSWATFTVKDLLMNDRLPGKKT--TKLWV 813
Db 2121 ekdeavnkiaslaeelkiltkemdefrdskes---lqeqsghiseelctyktelqmlkq 2176
Qy 814 PDEEVSPETVMYKIQAIKMMVWRLGLMKNN-----HSKSGTSTRLTLTLHSDGDLTEOGK 869
Db 2177 qkedinnklaevkvedellqhlslskeqlqigmeirneklr-----nyelcekmd 2228
Qy 870 ISKPDMSRLRLAAGSAIVKLQAQPCYHIEITLQYQICALAINDECYQVRO-----VFAQK 925
Db 2229 imekeisvrlr-----mqnepqgeeddvaermidile-srnqeiqlmekisavysseq 2279
Qy 926 LHKGLSLRLPLEYMAICALCAKDPVERRAHARQCLVKVNNVRREYLK----- 974
Db 2280 -htllsalselq-----keteah-khcm---lnikesstlsrsfsgslqt 2321
Qy 975 QHAASVEKLLSLEPYVVPYTIHLHADPDYVKVQDIEQLKRVKLECLFVLEILMAKNEN 1034
Db 2322 ehvkintqlqtlknkfkvvrytaavkedhslik--dye--kdlaa-----eq 2364
Qy 1035 NSHAFIRKWNENIKQTKDAQCPDDAK-----MNEKIYTVCDVAMNIIIMSKSTYS 1084
Db 2365 krhdelrlqlcleq-hgrkwsdsaseelkfcetefinelifk-----kanliqsvqddfs 2419
Qy 1085 -----LESPPK-----DPVLPAREFTOPDKNFSNTKNVLP 1114

Db 2420 evqvflnqvgtlgeelehkkqfmqwlfeefgdllhvdakklsqmqgqenrriastiqlitk 2479
Qy 1115 EMKGFF-TPGKPKTTNVLGAVNKPLSAGKQSQTKSSRMETVSNASS---SSNPSSPGRI 1170
Db 2480 rlkavvqskiqreiltvynlqfeaklqekqcnkelmrmehhgppsasvmeeenarllgil 2539
Qy 1171 KGRIDSSEMDS-----ENE-----DYTMS-----SPLPGKKS 1199
Db 2540 ktvqdesakklqsrklmlenenlnlvkddamhgekvaillqdkllsrnaeeelnamqvkitk 2599
Qy 1200 KRDSDDLVRSELE---KPRGKKTPTVEOEELKGMDDLTKLV--OEQPKGQSORSRKRGH 1254
Db 2600 kqdnlgamkeienlqkmvakgavpykeeiidnl-----ktkvkiemekikyskatdga 2655
Qy 1255 TASDESDEQWPEERKLKEDI-----LENEDEQNSPPKKGRGPPKPLGGGTPKKEPTMKT 1310
Db 2656 ylkscledkeegrlrkeelrraqadndttvcvpkdyqkastfvtcggsqgviqgstaml 2715
Qy 1311 SKGSKKSGPPAPPEEEEEERQSGNTEQKSKQHVRVRRRAQOARAESPESSATESTQS- 1369
Db 2716 vlq-----sekaalerelsh---ykkkyhhlsrtmsssedrktktaksdahss 2760
Qy 1370 -TPQKGRGRPSKTPS---PSQPKN 1390
Db 2761 htgsshrsgphkcteyrhgvpvtpers 2786
RESULT 14
AAG39233
ID AAG39233 standard; Protein; 1562 AA.
XX AAG39233;
XX XX
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 48514.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

625 glpessgsvkelqeenksklkeireresiektalie-----lemmeklvqknl 674
478 YLYATLDLNA-----VKALNEMKCONLLR-----HQVKDLDDLKQPKTDASVKA 524
675 llensisdnaeetirgikliee--asmslaeeksglhsekdmrlsrgatsens-kk 731
525 IFSKVMVITRNLPDGPKAQDFMKKTQVLED-----DEK-----IRKQ 562
732 lseemvlienslfanvealeeksklleeschnldkttltseresllshidtmkr 791
563 LEVLVPTCSCK-----QAEGCVREITKLG---NPKQPTNPFLEMIKFLLERI--- 608
792 ledlekehaelkvkvelateresslqki-eelgvslnakd--ceyasvfqsesrmngm 848
609 -APVHIDFESTALLKQVNSIDGTADDEGVPTDQAIRAGLELLKVLSTPHISFUSA 667
849 estihlhdengcrvreyqveldradhahleivlqkclqdwl----- 891
668 EFESILLACLKMDDEKVAEAAQIFKNTGSKTEEDFPHRSALLPVLHKKKGGPPROAK 727
892 eksssia-----engdikeas-klliekivseleee-----nigkvqvid 930
728 YAIHCIAFFSKETQFAQIFPL-----HKSLDPSN-----LEHLITPLVTI 770
931 ssincikil-----rtgiyvlmklelpgigsgdensrdqrmhldilrledmqtmllisi 986
771 ----GHIAL--LAPQFAAPKWSWATFIVKDLLMNDRLP----- 804
987 rdenqhsaienlileflrqkseavgletekkileeelesqqlsfsrdetqklifvn 1046
805 GKTKTKLWVPDEVSPEPMVKIQAIAKMMVRLGLMKNHNSKSGTSLRLTLTLHSDGQ- 863
1047 gelttkv---nggvnrekvlmve-tedfrhqvlgirday-----tilggdnk 1090
864 -ITEQKISKPDMSRLRLAAGSAIVKLAQPCYHEITLEQVQLCALAINDECYQ----- 917
1091 tidekayltk-----stlgleeekckle-----ddisllsetiyqsnlli 1131
918 -VROVFAQKLHLKLSRLRLPLEYMAICALCAKDPKERRAHARQCLVKNINVRREYLKQH 976
1132 lledvilekls-sgamklnedldrlsvkckleeevreygdklsadianfqivgleksn 1190
977 AAVSEKLLSLPEYVYPTIHLHLADPDVVKVQDIEQLKDVKEKFWLFIEMKANNENS 1036
1191 a----ellsarsa-----nvh-leheianvkq-----kekelleamlmsimneke 1234
1037 HAFIRKMWENI-----KQTK-----DAQGPDADAKMNEKLYTVCDAVMI 1075
1235 ---lskaveglecrykeakeaieedrkgvlrlrgdydeqvkknshsneanlkleadlml 1291
1076 IMSKSTYSLESFK--DPVLPARFTQDPKNSNTKYNLPP-EMKSFTPGPKPTNV-- 1130
1292 lm-----eleekvkekenlnqelfte-----rnelelwesqsatlfgelqisavhe 1337
1131 --LGAVNKPPLSAGKOSQTKSS-----RMETVSNASSSN-----PSSPGRIK 1171
1338 tllegltneiveacknleersitlkareqlkgrvnniedankgndlmckyqaifilk 1397
1172 GRLDSE-----MDHSENEYDTMSSPLPG-----KSKDRDSDSLVRSELEKPRGRKTPVT 1223
1398 esiqslskehamihenfengpattngsfvgsyqetaslvdsngdf-leiqelhirikaeee 1456
1224 EQEELGMDDLTKLVQEQPKQSQRKRGHTASTESDEQOWPEERKL----- 1270
1457 aitkklameel-----ktsaarsrrngslrkqnheiyseeteemtkdivldqvsdc 1509
1271 -----KEDILENEONSPPKKGRGRPPKPLGGTTPKEPTMTSKSGSKKSGPPAP 1324
1510 ssygistrdilkiedhs---leaksqppk---gkslseeeslvvdkleisdrftdpnk 1562
1325 EEEEEERQSGNTEQKSKQHVRERRAQRAESPSSAISTQSTPQKGRGRPSKT 1381
1563 dankkvlerlnsdliqklnlhvavedlikve-----teekdekkeneyet 1610

RESULT 13

AAY01632
ID AAY01632 standard; Protein; 2954 AA.

AC AAY01632;

DT 22-JUN-1999 (first entry)

DE Amino acid sequence of centromere-associated protein-E (CENP-E).

KW CENP-E; centromere-associated protein-E; ATPase activity;

KW plus end-directed microtubule motor activity; chromosome congression;

KW microtubule binding activity; chromosome movement; mitosis;

KW cell proliferation; tumor; metastasis; vascular malfunction;

KW inflammatory disease; immune disease; angiogenesis; hypertension;

KW restenosis; fungal infection; selective herbicide; fungicide;

KW insecticide; plant growth regulator; activator; cancer cell marker.

OS Xenopus sp.

XX WO9913061-A1.

XX 18-MAR-1999.

XX 10-SEP-1998; 98WO-US19231.

XX 11-SEP-1997; 97US-0058645.

XX (REGC) UNIV CALIFORNIA.

XX Cleveland DW, Goldstein LSB, Sakowicz R, Wood KW;

XX WPI; 1999-229233/19.

XX N-PSDB; AAX26819.

XX Centromere-associated protein-E and related nucleic acid

XX Claim 5; Page 66-67; 77pp; English.

The present sequence represents CENP-E (centromere-associated protein-E) of Xenopus. The protein has at least one of plus end-directed microtubule motor activity, ATPase (adenosine triphosphatase) activity and microtubule binding activity. CENP-E is the motor that powers chromosome movement toward microtubule plus ends and is essential for congression of chromosomes during mitosis. Modulators of CENP-E can thus control cell proliferation. Agents that modulate CENP-E activity are lead therapeutic, bioagricultural and diagnostic agents, e.g. for treatment of unwanted cell proliferation (typical of many examples are tumors and metastases; vascular malfunction; inflammatory and immune diseases; angiogenesis; hypertension; restenosis; and fungal infections); also as plant-protection agents (selective herbicides, fungicides and insecticides) and plant growth regulators or activators for improving yields. CENP-E is also a diagnostic marker for dividing cells, including cancer cells.

SQ Sequence 2954 AA;

Query Match

Best Local Similarity 2.7%; Score 192; DB 20; Length 2954;

Matches 294; Conservative 239; Mismatches 575; Indels 478; Gaps 69;

QY 110 ITRQLKGLDET---KSPQFNRYFYLLLENIAWVKSYNICFELEDSEIFQYTLFSVIN 166

Db 1374 lckkingleleikgseeevlksmlenlk-----ednnklkeqa-----eys 1417

QY 167 NGHNQKVHMHVMDLMSIIICEGDTVSQELLDVNLVPAHKNLNKQAYDLAKALKRPA 226

Db 1418 skeng---fsleevfsg-----sqklvdei-----evikaqikaee 1450

QY 227 QAIEPVIITFF-----NQVLMGLGTSISDLSEHVDLI----- 259

897 ssincikil-----rtgiyvImkleiIpiqsgdensrdgrnmhdilnrlmedmtllisi 952
771 ---GHIAL--LAPDOFAAPKSWVATFIVKDLLMNDRLP----- 804
953 rdenqhsaenlvleflrlqkseavgietekkleesqcgqlsfrdetqklifvn 1012
805 GKTKTKLWPDDEEYSPETMWKIQATKMMVRWLLGMMKNHKSSTGLRLLTILHSDGD- 863
1013 geltkv---nggvnrekvlave-ledfhrqvlrlday-----tllqgdunk 1056
864 -LTEGKTSKPDMSKRLRLAAGSAIVKLAQEPYHEIITLEQYQCALAINDECYQ----- 917
1057 tldkayitk-----stqlkeekckle-----ddislllsetiyqsnlii 1097
918 -VRQVFAOKLHGKLSRLRLPLEYMAICALCAKDPKERRAHARQCLVKNINVRREYLKQH 976
1098 lledvilekl-sgamklnedlrisivckleeevregldkksaadianfqlqvpleksn 1156
977 AAVSEKLLSLLPEYVVPYTHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNS 1036
1157 a-----ellisarsa-----nvh-lehianvkvq-----kekelleamlmisinqekse 1200
1037 HAFIRKMYENI-----KQTK-----DAQGPDDAKMMNEKLYTVCDAVAMI 1075
1201 ---lskaveglecrykeakaieedrkvrlrgdydeqvkknshsneanikleadlmln 1257
1076 IMSKSTTVSLSPK--DPVLPARFTQDPKNFNTKNVLP-EMKSFPTPKPKTTNV-- 1130
1258 lm-----eleekvekenlnqelfte-----rneielwesqsatlfgelqisavhe 1303
1131 --LGAVNKPLSSAGKOSOTKSS-----RMETVSNASSSSN-----PSPGRIK 1171
1304 ttleqthelvecaknleestlkdreieqlkgrvnnledankgndlmckyagaifilk 1363
1172 GRLDSE-----MDHSENEYTMSSPLPG-----KKSDDRDDSLVRSELEKPRGRKKTPTV 1223
1364 esiglekhamlhefengpattnqsfvgsygetaslvdsndgfl-leiqelhlrikaee 1422
1224 EOEELGHDDTLKLVQEQPKGSQSRKRGHTASDESQOWPEKRL----- 1270
1423 altkklameel-----ktsarrsrringslrkqnheiyseetemitkddivldqvscd 1475
1271 -----KEDILENEDEQNSPPKGRGRPPKPLGGGTPEKEPTMTKSKGSKKSGPPAP 1324
1476 ssyglstrdilkiedhs---leaksnppk---gkslseeslvvdkleisdrftdpnk 1528
1325 EEEEEERQSGNTEQKSKOHRVSRRAQRAESPESAIESTOSTPOKGRGRPSKT 1381
1529 dankrvlierinsdlqklsnhvavedikikve-----teekdekkeneyet 1576

RESULT 12
AAG36712
ID AAG36712 standard; Protein: 1744 AA.
XX
AC AAG36712;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45030.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX

25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126284.
PR 25-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

1194 ---lskaveglecrykeakeaieedrdkgvrlrlrgdydeqvkknshsneanlkleadlml 1250
1076 IMSKSTTYSLESPK--DPVLPAFFFTOPDKNFNTKNVLP--EMKSFTTPGKPKTTNV-- 1130
1251 lm-----eleeikvekenlqelife-----rnelwesdaatlfgelqisavhe 1296
1131 --LGAVNKPSSAGQSOTKSS-----RMETVSNASSSSN-----PSSPGRIK 1171
1297 tllegltnelvaeacknlestrtkdreieqIkgvnniedankgndlmcyaqaifilk 1356
1172 GRLDSE-----MDHSENEYDYMSSPLPG-----KSDKKRDDSLVLRSELEKPRGRKKTPTV 1223
1357 esiqslekhamlhfengpattngsfvqysetsasldvnsdgf-leiqelhlrikaatee 1415
1224 EQEELGMDLTKLVQEQPKGSORSRKRGHTASESDQQWPEEKRL----- 1270
1416 aitkklaameel-----ktssarrsrrngslrkqnheiyseetemitkdilvdqvsdc 1468
1271 -----KEDILENEDEQNSPPKKKGRGRPPKPLGGTPEKEPTMTKSKGKKSGPPAP 1324
1469 ssygistrdilkiedhs---leaksnppk---gkslseeislvdkleisdrftdpnk 1521
1325 EEEEEERQSGNTEQKSKQHRVSRRAQRAESPSSAISTQSTPQKGRGRPSKT 1381
1522 dankkvlerlnsdIqklsnlhvavedlkikve-----teekdekqkeneyet 1569

RESULT 11
AAG36713
ID AAG36713 standard; Protein; 1710 AA.
AC AAG36713;
XX
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45031.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
XX
PL 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

2426 sgtilrepttksvpvnlpersptdspre 2454

RESULT 10

AAG36714

AAG36714 standard; Protein; 1703 AA.

AC AAG36714;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 45032.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
04-MAY-1999; 99US-0132407.
05-MAY-1999; 99US-0132484.
06-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
07-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.

18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.

24-OCT-1994; 94US-0328254.
22-OCT-1993; 93US-0141239.
(TEXA) UNIV TEXAS SYSTEM.

Lee W, Zhu X;

WPI; 1998-109817/10.
N-PSDB; AAV09076.

New isolated mitosis protein and gene - useful for, e.g. developing products for therapy and diagnosis of hyper-proliferative disorders such as cancers or psoriasis

Claim 1; Column 40-52; 43pp; English.

This is the amino acid sequence for mitosis, a phosphoprotein necessary for the cell to enter mitosis. The protein's degradation is necessary for the cell to advance into the next stages of mitosis. The mitosis protein, can be used to control the growth of cells. An anti-mitosis antibody, a mutant or a non-functional analogue of mitosis can inhibit the mitotic cell cycle by preventing the cells from entering the M phase, and over expression of mitosis or its functional equivalent, would inhibit the cycle by preventing cells from leaving the M phase. Antagonists to this protein can be used to control hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease, psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast cancer, sarcomas and other neoplasms, bladder cancer, colon cancer, lung cancer and various leukemias and lymphomas). Reintroduction or supplementation of lost mitosis function by introduction of the protein or nucleic acid encoding the protein into a cell can restore defective chromosome segregation, which is a marker of progressing malignancy. Malignant proliferation of cells can then be halted. The protein can also be used for the detection and diagnosis of hyperproliferative cells.

Sequence 2482 AA;

Query Match 2.7%; Score 194.5; DB 19; Length 2482;

Best Local Similarity 19.4%; Pred. No. 0.00058;

Matches 297; Conservative 233; Mismatches 592; Indels 407; Gaps 63;

19 VKESDKI-----SKEEMVRLKMYVTFMDMDQDSEEEKLY-----56

1172 mkeidskhlhqvqlmkieacielekvigeklenskdslekleyfscdqhellqrvtvs 1231

57 --LNLALHLSDFELKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLXDFMFTROL 114

1232 eglnsdlembad---ksredigdnva-----kvndswkerfldvenel 1272

115 KGLDRTKSPQFNRYFLLNIAWVKSYNICFELEDSN--EIFTQLYRTLFSVINNHGK 172

1273 srirsekastehealyleadlevvqtcklekdnenkqvicleeel-svtsernql 1331

173 VHMHMVDMLSIIEGTVSOELLDTVLVNLVPAHKNLNQAYDLAKALKRRTAQAEIPEY 232

1332 --rgeldtmskttaldqlsekemek--tqeshqshclicqvaaevkektellq-- 1385

233 ITTFFNQVLMGKTSISDLSHFVFDLI-----LELYNIDSHLLSVLPOLFELKLSNDNE 287

1386 -----Cissdseilkkdkhlqekqlskdsqalsltkelengiaqlnke 1432

288 ERLQV--VKLIAMFGARDELSAQNKPLMOCYL--GRFNDIHPVIRLECVKFASHCLMN 343

1433 kelliakesesiqarlssdesyeklnvskalealvekegef-----alrlsstqeevhqlr 1487

344 HPDLAKDLTEYLVKVRSHDPEAIRHDVIVSVTAACKDILLVNDHLLNFRVREPLDK--- 400

1488 g-----ieklrvriadekqh--iaeklkererendslkdkvenlerelqmseenq 1538

401 -----RWRVRKEAMWGLAQIYKKYALQ-----SAAGRDAQAQIAWIKDKLLHIY 444

Db 1539 elvildaenskaevetlktqieemarslkvfeldvlvtrsekenltkqikeqgglseld 1598

QY 445 YQNSIDDRLLVERIFAQYMPVPHNLETTERRMKCLYLYATLDLNAVAKALNEMKQCNLLRH 504

Db 1599 klissfksllcekeqaeiqikeesktavem-----lqn 1631

QY 505 QVKDLDLIKOPTDASVKAIFSVMVITRNLPDP-----GKAQDFMKFFQVLEDDDKI 559

Db 1632 qikelneavaalcgd-----qeimkatqegldppieehqlrnsieklrleadek- 1683

QY 560 RKQLEVLVSPTCSCKQAEGCVREITTKLGNPKQPTNPFLEMIKELLERAPIVHDITESIS 619

Db 1684 -kgicvliq---qikesehadlikgrvenlere---lelartnqehaa---leaeensk 1731

QY 620 ALIKQVKNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSTFHPISFSAETFESILLACLKM 679

Db 1732 gevetlkakiegmt-----qslr-glel-----dvtirsek--enltneqlk 1771

QY 680 DDEKVAEAL--QIFKWTGSKIEEDFPHIR---SALLPVLHHSKSGKPPROAKYAIHCIIH 734

Db 1772 egeriselelinssfenilqekeqevgmkeksstamemlqtqikeinerva--alhdndq 1829

QY 735 AIFSSKETQQAIFEPHLH-----KSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKS 787

Db 1830 eackakeqnlssqvceclelekaqllgldakennylviqssvng----- 1873

QY 788 WYATFIVKDLMLNDRLPKKTKTLWVPDEVSPTMTVQIAQKMMRWLLCMKNHNS--K 845

Db 1874 -----liqevedgk--klekdeels-rlknqiddqevklsgvegehqwlk 1920

QY 846 SGTSTLRLTLILHSDGLTEQGIKISKPDMSRLRLAAGSAIVKLAQEPYHEITLBOYQ 905

Db 1921 eqnlelnltveleqkivqisknaslqg-----tlevlq 1955

QY 906 LCALAINDECYQV---QVFAQKLHKLGLSR---LRLPLEYMAICALCAKDPV---KERRA 956

Db 1956 ssyknleleleltkmdkmsrvekvnkmtaketelqremhemaqktaelqeelsgeknrla 2015

QY 957 HARQCLVKNINVRREYLKQHAHVSEKLLSLLP-----EYVVPYTHLLAHD 1002

Db 2016 gelqlleeksskdqkeltlensekksldcmhkdqvekegkvreeiaeyqirl--he 2073

QY 1003 PD-----YVVKQDI--EQLKDVKECL---WVLEILMAKNE--NNSHAFIRK 1042

Db 2074 aekhqallidtnkqyeiqtyrekitkskeclssgkleldllkskeelnnslkattq 2133

QY 1043 MVENIKOTKDAQGDDAKMNEKLYTVCDVA---NMIIMSKSTTVSLESPPKDPV-----L 1093

Db 2134 ileelkktk---mdnlkyvnnqlkkeneraqgkmlili--ksckqleeekeellqelsql 2187

QY 1094 PARFFTQPKDNFSNTR--NYLPPEMKSFPTPGKPKTTWV-----LGAVNKPLSSAGK 1143

Db 2188 qaaqekktgtvmdtkvdeitlkelketleektkeadeylkycsllishekleake 2247

QY 1144 QSQTQ-----SSRMETVSNASSSNPSPGRIKRLDSSEMDHSENEDYTHSSPLP----- 1194

Db 2248 mletqvahlcsqgskdsgspilgvpvg-----pspipsvte 2286

QY 1195 -----GKSKDRDSDSLVRSELEKPRGKKTPTVTEQEEKLGMDLLTKLVQEQKPKGSQ 1247

Db 2287 krissgngnksgrkq-----rsggiwengggtpat-----pesfskksk 2326

QY 1248 RSRKRG--HTASEDEQOQWPEEKRLKEDILENEDQNSPPKGRKGRPPKPLGGGTP---KE 1304

Db 2327 kavmsghpa-----edtegtefepegilpevvkkgfadiptgktsyillr 2371

QY 1305 EPTWKT---SKKGSKKSGPPAPPEEEEEERQSCNTQKSKSKOHRVSRRAQQAESPES 1361

Db 2372 rttmatrtsprlaaqklalsplslglenlaesskptaggsrsgkvkvaqr-----spvd 2425

QY 1362 SAI-----ESTQSTPQKGRGRPSKTPSPSQ 1386

2216 g-----ieklrvieadekkqlh--iaeklkererendsalkdkvenlerelqmseenq 2266
402 -----WVRKEAMMGLAOIYKVALOSAAGKD-AAKOIAWIKOKLLHIY 444
2267 elvildaeenskavetiktgleenarskifeldvltirsekentlqkqegqqlseid 2326
445 YONSIDRLVVERIFAQYVMVPHNLETTERMKCLYLYATLDLNAVAKALNEMKCONLLRH 504
2327 kllssfskllseekaeiqikeestavem-----lqn 2359
505 QVKDLILLIKOPKPTDASVAKAIFSKVMVITRNLPD-----GKAQDFMKKFTQVLEDEDEKI 559
2360 qikelneavaalcgd-----qelmatedslpdpieehqrlnssteklirarleadek 2411
560 RKQLEVLVPTSCSCQAGCGVREITTKLGNPKQPTNPLEMIKFLLEIRIAPVHIDTESIS 619
2412 -kqlcvq----qkesehhdllkgrvenlere----leartngqehaa---leaensk 2459
620 ALIKOVNKSIDGTADDEGEVPTDOATRAGLELLKVLSTFHPISFHSAAETEFESLLACLKM 679
2460 geveltkaklegmt-----qslr-giel-----dvvrtirsek--enltneqlq 2499
680 DDERVAEAL--QJFKNTGSKIEEDFPHIR---SALLPVLHHSKSGPPROAKYAIHCIIH 734
2500 egeriseleifnsfennllqekedqvmqeksstamemlqtqikelneriva--alhnndq 2557
735 AIFSSKETQFAIPEPLH-----KSLDPSNLEHLITPLVTIGHIALLPDQFAAPWKS 787
2558 eachakeqnissqvecieleakqllqgldeaknyivlqssvkg----- 2601
788 WATFIVKDLMLNDRLPGKTKTKLVPDEEVSPTVMYKIAIKMVRWLLGMKNHNS--K 845
2602 -----llqevdgkg--klekdeels-rlknqigqegqlvsklqvegehgqlwk 2648
846 SGTSTLRLLTILHSDGLTEQKISRPDMRSRLRLAAGSAIVKLAQPCYHEIITLEOYO 905
2649 eqnlelmitveleqkqiglsknaqlq-----tlevlq 2683
906 LCAIAINDECYQV---ROVFAQKLHKLGLSR---LRLPLEYMAICALKADPV---KERRA 956
2684 ssyknlenelctkmdkmsfvekvnmktaketelqremhemaqtaelqeelsqeknrla 2743
957 HAROCLVKNINVRREYLKQHAANVSEKLLSLP-----EVVVPYTHLHAHD 1002
2744 gelqllleekskdqkeltlenselskldcmhkdqvekgvreesiaeyqirl--he 2801
1003 PD-----YVKVODI--EOLKDVKECL--WVLEILMAKNE--NNSHAFIRK 1042
2802 aekhqallldtnkqyveidqyrekitskeecissqkleidilkskeelnnslkatq 2861
1043 MVENIKOTKDAQDPDARKMNEKLYTVCDA---MNIIMSKSTTYSLESKDPV-----L 1093
2862 ileelkktk---mdnlkynqlkkeneragkmlli--ksckgleeekelqkelsql 2915
1094 PARFTQPKNFSWK--NYPPEKNKSFPTPKPKTTNV-----LGAVNKLPLSSAGK 1143
2916 qaaqekqktgtvmdtkvdelteikelteektkeadeyldkycsllishekleake 2975
1144 QSOTK-----SSRMETVSNASSNSPSPGRIKGRLDSEMDHSENEDYTMSSPLP---- 1194
2976 mletqvalhsqgqkdsrgslpgvpvgp-----pspipsvte 3014
1195 -----GKSDKDDSDSLVSELEKPRGRKKTPTVEQEEKLGMDLTKLVOEQKPGSQ 1247
3015 krlssgqnkaagkq-----rsgsiwengrtpat-----pesfskksk 3054
1248 RSRKRG--HTASESDQWPEEKRLKEDILENEDEONSPPKGKRGPRPKPLGGTTP--KE 1304
3055 kavmsghpa-----edtegtefepglpevvvkgfadiptgktspyllr 3099
1305 EPTMKT---SKGSKKSGPPAPPEEEEEERQSGNTEQKSKQHRVSRRAQRAESPES 1361
3100 rtmatrpsrlaqaikalsplsigkenlaesskptaggsrskvkaqr-----spvd 3153

OY 1362 SAI-----ESTQSTPQKGRGRPSKTPSPSQ 1386
Db 3154 sgtlrepttksvpnnlpersptdspre 3182

RESULT 8

AAR72826
ID AAR72826 standard; Protein; 2482 AA.
XX AAR72826;
AC AAR72826;
XX 27-FEB-1996 (first entry)
XX Human mitosis.
DE Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth;
KW inhibition.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Region 1480..1659
FT /label= internal_repeat
FT Region 1660..1839
FT /label= internal_repeat
XX WO9511309-A2.
XX 27-APR-1995.
XX 24-OCT-1994; 94WO-US12162.
XX 22-OCT-1993; 93US-0141239.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Lee W, Zhu X;
XX WPI; 1995-170229/22.
XX N-PSDB; AAQ86851.
XX Purified mammalian protein mitosis and agents that bind it and
PT inhibit its action - used to promote cell growth or to inhibit cell
PT division and/or proliferation
XX Claim 4; Fig 8B; 61pp; English.
XX AAR72829 is human mitosis. Mitosis is involved in the regulation of
CC the mammalian mitotic cell cycle. Mitosis as with E2F-1 (see AAR72824)
CC interacts with the retinoblastoma protein (the retinoblastoma tumour
CC suppressor gene product). Mitosis is first synthesised at the G1/S
CC boundary, it is then phosphorylated from S through M phase, and during
CC mitosis, is closely associated with the centromeres/kinetochores at the
CC mitotic spindle poles. Mitosis is necessary for a eukaryotic cell to
CC enter the M phase of the mitotic cell cycle and its degradation is
CC necessary for a cell to advance on to the next stage. Mitosis is thus
CC useful for controlling cell growth as overexpression of mitosis prevents
CC a cell from exiting the M phase.
CC An anti-mitosis antibody, antibody fragment or a phosphorylated mitosis
CC muten (or nucleic acid encoding it) can also be used to inhibit cell
CC division which is particularly useful for the study of the cell cycle.
CC A further use is to control hyperproliferative cells, and so control
CC diseases such as psoriasis and breast cancer. It can also be used to
CC block gametogenesis of an immature gamete.
XX SQ Sequence 2482 AA;

Query Match 2.78; Score 194.5; DB 16; Length 2482;
Best Local Similarity 19.48; Pred. No. 0.00058;
Matches 297; Conservative 233; Mismatches 592; Indels 407; Gaps 63;

obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures - Claim 13; SEQ ID 6892; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 92 AA;

Query Match 6.5%; Score 467; DB 21; Length 92;
Best Local Similarity 100.0%; Pred. No. 5.5e-24;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

839 MNKHSKSTSLRLTLTLHSDGDLTEGKISKPDMSRLRLAAGSAIVKLAQEPYHEI 898

1 mknhsksstslrltltlhsgdldteggkiskpdmrlrlaagsaivklaqepcyhei 60

899 ITLEQYQICALAINDECYQVROVFAQLHKH 929

itleqyqicalaindecyqvrqvfaqlhkg 91

RESULT 6

AAW98879
AAW98879 standard; Protein; 1743 AA.

AAW98879;

31-MAR-1999 (first entry)

H. pylori GHPO 1755 protein.

GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis; peptic ulcer disease.

Helicobacter pylori.

WO9843478-A1.

08-OCT-1998.

01-APR-1998; 98WO-US06371.

29-JUL-1997; 97US-0902615.

01-APR-1997; 97US-0833457.

24-JUN-1997; 97US-0881227.

(HUMAN) HUMAN GENOME SCI INC.

(INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

WPI; 1998-542293/46.

N-PSDB; AAX14598.

New isolated Helicobacter polynucleotides - used to develop products

for the diagnosis, prevention and treatment of Helicobacter

infections and gastrointestinal diseases

Claim 8; Page 2000-2008; 2054pp; English.

XX

This sequence represents a Helicobacter pylori GHPO protein of the invention. The polypeptides can be used for preventing or treating Helicobacter infections, and gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

Sequence 1743 AA;

Query Match 3.1%; Score 225; DB 19; Length 1743;
Best Local Similarity 18.4%; Pred. No. 3.4e-06;
Matches 285; Conservative 217; Mismatches 523; Indels 520; Gaps 71;

QY 4 SKTRTN-DGKITYPGCVKEISDKISKEEMVRRLK-----MVRVT 41

Db 312 tetqtnfgdks-----eeitddndqelksgkkyiigivavlvilfsisfhy 366

QY 42 FMDMDQDS---EEKELYLNLAHLASDF--FLKHPGKDVRLVAVCLADIFRIYAPEAP 96

Db 367 fmpledksrfskdrnlvndeigirqeynrllkernekgnmid----- 410

QY 97 YTSPPKLDIFMFIYRQLKGLDTSKPOFNRYFYLLNTAWKSYNICPELDSNEIFTQ 156

Db 411 -----knlf-----nddpnrtlyn1-nla-----eiedknpl----- 438

QY 157 LYRTLFSVINNGHOKVHMHVMDLMSIIICEGDTVSQELLDTVLVNLPVPAHKNLKNQAVD 216

Db 439 --rafyecisngngyeeciklli-----kdkklqdg----- 466

QY 217 LAKALLKRTAQAIIEPYITTFNQVLMGLKTSISLSEHFVLDLILELYNIDSHLLSLVLPQ 276

Db 467 -----mkktlea-----yndciknakt----- 483

QY 277 LEFKLSNDNEERLOVVKLLAKMFGAKDSELASQNKPLMQCYLGRFNDIHVPIRLECVKF 336

Db 484 -----eeerikcldli-----kdenl-----lngqkvqvaldcikn 518

QY 337 A-----SHC---LMNHP-----DLAKDLTYLVKVRSHDPEAIRHVDIVSIVTAA 378

Db 519 aktdeerneckilndpeirekfrkelelqelqey-----kdciknaktea 565

QY 379 KKDILL--VNDHLLNFVREITLD-----KRWVRKEAMMGLAQIYKKYALQ----- 422

Db 566 ekncklqiskeaierlkqaldckknakteernecklnipqdlqkelladmsvskaykd 625

QY 423 --SAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLIVERIFAQYVMPVPHNLET--TERMKCLYY 479

Db 626 cvskarnekeqec---eklltpearkklegqv-----dciknaktdeerkcklkd 674

QY 480 LYATL--DLNAVKNL-----EMWKNCONLL-----RHQVKDLDDL 512

Db 675 lpkdlqsdilakeslkaydcvskaqteakeckeklltpeakilleeakesvskayldc 734

QY 513 IKQPKTDASVKAIFSKVMVITRNLPDPGKAQDMKFKFTQVLED-----DEKIRKQLEVL 566

Db 735 vsqakteakekce-ekllt-----peakkkleekskayldcvskarnekekecekl 788

QY 567 VSPTCSCKQAECVREITKLGKNGKQPTNPFLMTKFLLERIAPVHIIDTESISALIKQVN 626

Db 789 ltpeak-killeqgaicdknaktkerkckldlpkdlqkv-----lakesvskayldcv 843

QY 627 KSIDCTADDEGVPTDQAIRAGLELLKVLSTPHISFSAETFEESLACLAKMDDEKVAE 686

Db 844 qakteakekecklltpea-rkllleak-----ksvkaxidc-----vsq 882

QY 687 AALQIFKWTGSKIIEEDFPPIRSALLPVLHHSKKGPPRQAKYAI-----HCIHAIFSSKET 742

Db 883 akteakekeckllt--pearkll-----eeakesvskaydcvskarneke 927

QY 743 QFAQIFELPHKSLDPSNLEHLITPLVTIGHTIALLAPDQFAAPWKSVAFTIVKDLLMNDR 802

Region /note= "Hank's conserved region - beta strand 4"
525..548
/label= "Subdomain V"
Region /note= "Hank's conserved region"
525..532
/label= "Beta strand 5"
Region /note= "Al[ha helix D"
540..548
/label= "Al[ha helix D"
Region /note= "Subdomain V1a"
554..567
/note= "Hank's conserved region - alpha helix E"

W0200050454-AL.

31-AUG-2000.

24-FEB-2000; 2000WO-US04732.

24-FEB-1999; 99US-0121461.

(TUFT) TUFTS COLLEGE.

Soto AM, Sonnenschein C, Geck P, Szelei J;

WPI; 2000-565451/52.

N-PSDB; AAA28051, AAA28052.

New human androgen-induced tumor suppressor cDNA sequence termed

'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient

diagnosis and treatment of prostate cancer -

Claim 2; Fig 1; 152pp; English.

This invention relates to a human androgen-induced tumour suppressor cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell proliferation and use as a marker for the efficient diagnosis and treatment of prostate cancer. The invention includes AS3 cDNA and protein sequences, a vector comprising the cDNA sequence, a host cell transfected with the expression vector, and a method for producing an AS3 polypeptide comprising culturing the transfected cells. AS3 has cytostatic activity, and acts to suppress cell proliferation. The AS3 gene is useful as a marker for the efficient diagnosis and treatment of prostate cancer. The AS3 nucleic acid molecule can be used as a source of antisense agents for sequence specific modulation of gene expression. The AS3 protein may be used in the treatment of disorders caused by aberrant modification or mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene or aberrant post-translational modification of the AS3 protein. This sequence represents the human AS3 protein sequence.

Sequence 1391 AA;

Query Match 100.0%; Score 7193; DB 21; Length 1391;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAHSKTRTNDGRTYPPGKVEISDKISKEVMRRLKVVVKTFFMDQDQDSEEEKELYNLIA 60
|||||
1 mahsktrtndgkityppgkveisdskiskeemvrrlkmvvtffmdmdqsdseeekelynlia 60
|||||

61 LHLASDFFLKHGKQVRLVACCLADIFRIYAPEAPYSPDKLIDFIMPITRQLKGLED 120
|||||
61 lhlasdfflkhpqkdvrlvavccldadifriyapeapyspdklklidfmftrqlkgled 120
|||||

121 KSPQNRXYLLNTAWKSYNICFELEDSNEIFLTOLYRTLFVNLNNGHNOKVHMHVVDL 180
|||||
121 kspqnfryllylntawksynicfeleedsneifltylrvtnlsvnngnkvhmvhvd 180
|||||

181 MSSIICEGTVSQELLDVTLVNLVPAHKNLNKQAYDLAKALKRRTAQAIPEYITTFNFQV 240
|||||
181 mssicegdtvsqellldvtlvnlvpahknlnkqaydlalakalkrtaqaiepyitffnqv 240
|||||

QY 241 LMLGTSISDISEHVPDLILELYNIDSHLLSVLPOLFEK LKASNDNEERLQVVKLAKMF 300
|||||
DB 241 lmlgtsisdiselshvdfdlilelynidshllsvlpqlfeklksndneerlqvkvllakmf 300
|||||

QY 301 GAKDSELASQNKPLWQCYLGRFNTHVIRLCEVFASFHCLMNHDPDLAKDLEYLKVRS 360
|||||
DB 301 gakselasqnkplwqcygrfnthvirlcevfashclmmhdpdlakdleylkvrs 360
|||||

QY 361 DPEEAIHRHDVIVSIVTAARKOILLNDFNRERTLDKRWVRKEAMMGLAQIYKYA 420
|||||
DB 361 dpeearhdvsvivtaarkoillnfnvrtldkrwvrkeammglaiqykya 420
|||||

QY 421 LOSAAGKDAAKIAWKDKLLHIYYQNSIDDRLLYVERIFAQYVWPHNLETTTERMKCLLYL 480
|||||
DB 421 lqsaagkdaakiawkdkllhiyyqnsidrrllverifagymvphnlettermcklyl 480
|||||

QY 481 YATLDNAVKALENMVKONLIRHOVKDLIDIKPKPTDASVKAIFSKVMVITRNLDPDG 540
|||||
DB 481 yatldnavkalenmwkcnllrhqvkdldikpktdasvkaifskvmvitrnlpdpg 540
|||||

QY 541 KAADFMKKFTQVLEDEKIRKOLEVLVSPSCCKOAGCVREITKKLGNPKQPTNPFLM 600
|||||
DB 541 kaqdfmkkftqvleddekirkolevlvsptcsckgaegcvreitkklgnpkqptnpflm 600
|||||

QY 601 IKFILLERTAPVHIDTESISALIKQVKNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSTFH 660
|||||
DB 601 ikfllertapvhidtesisalikqvnksldgtadedegvptdqairagllellkvlstfh 660
|||||

QY 661 PISFHSATFESLLACLKMDDEKVAEALQIFKNKSGKTEEDFPHRSALLPVLHHKSK 720
|||||
DB 661 pisfhsatfesllaclkmddekvaeealqifknsgkteeedfphrsallpvlhhsksk 720
|||||

QY 721 GPPROAKYAIHCHAIFFSKETQFAQIFELPKSLDPSNLEHLITPLVTIGHIALLAPQ 780
|||||
DB 721 gpprqakyaiahchaihffsketqfaqifeplksldpsnlehlitplvtighiallapq 780
|||||

QY 781 PAAPKWSVATFIVKDLLMNDRLPKKTKLWVDEEVSPTWVKIQAIAKMMVRWLLGK 840
|||||
DB 781 faapkwsvatfivkdllmndrlpgkttklwvdeevsptwvkiqaiaikmmvrwllgk 840
|||||

QY 841 NNHSGSTSTRLTLTILHSDGDLTEQGIKSPDMSRLAAGSAIVKLAQBPCEHEIT 900
|||||
DB 841 nnhsgststlrltlilhsdgdlt eqgikspdmrslraagsaivk laqbpceyheitt 900
|||||

QY 901 LEQYQCALAINDECYQVRQVFAQLKGLSLRLPLLEYMAICCAKDPVERRAHARQ 960
|||||
DB 901 leqyqlcalaindecyqvrqvfaglkglslrlpleymaicalcakdpvkerraharq 960
|||||

QY 961 CLVKNINVRREYLKQHAASEKLLSLPEYVVPYTHLLAHPDVYKQDIEQLKDVKEC 1020
|||||
DB 961 clvkninvrreylkhaasekllslpeyvvpvpythllahdpdyvkqdieqlkdvkec 1020
|||||

QY 1021 LWFVLEILMAKNENNSHAFIRKVENIKQTKDAQGDDAKMNEKLYTVCDAVNNIIMSKS 1080
|||||
DB 1021 lwfvleilmaknennshafirkmveniktkdaqgddakmnekytvcdavnniimsk 1080
|||||

QY 1081 TTYSLSPKDPVLPARFFTPQDKNFNTKNYLPPENKSEFTPGKPKTKTNVLGAVNKLPS 1140
|||||
DB 1081 ttypslspkdpvlparrfftpqdknfnntknylppenkseftpgkpktktnvlgavnkplss 1140
|||||

QY 1141 AGKQSQTSRSMETVSNASSNSPSPGKGLDSESDHSENEEDYTMSSPLPGKSKDK 1200
|||||
DB 1141 agkqsqtsrsmetvsnassnspspgrikglldssedhsenedytmssplpgkssdk 1200
|||||

QY 1201 RDSDLVSELEKPRGRKKTPTVTEQEKLGMDLTKLVQEQPKGSQRKRKRGHTASESD 1260
|||||
DB 1201 rdsdlvselekprgrkktptvteqeklgmdltklvqeqpkgsqrkrkrghtasesd 1260
|||||

QY 1261 EQQWPEEKRLKEDILENEDEQNSPKKGRGRPPKPLGGGTGKEPTMTKSKGSKKSG 1320
|||||
DB 1261 eqqwpEEKRLKEDILENEDEQNSPKKGRGRPPKPLGGGTGKEPTMTKSKGSKKSG 1320
|||||

QY 1321 PPAPEEEEERQSGNTEQKSKSHRVSRRQAQRAESPESSAIESTQSTPQKGRGRPSK 1380
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:00:29 ; Search time 27.45 seconds
(without alignments)
3072.057 Million cell updates/sec

Title: US-09-512-581-2

Perfect score: 7193

Sequence: 1 MAHSKTRTNDGKITPPGVK.....QKGRPSKTPSPQPKNV 1391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 6062398 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7193	100.0	1391	21 AAY94702	Human androgen shu
2	2715.5	37.8	854	21 AAB01382	Neuron-associated
3	758.5	10.5	363	21 AAB58462	Lung cancer associ
4	513	7.1	101	21 AAY65247	Human 5' EST relat
5	467	6.5	92	21 AAG02811	Human secreted pro
6	225	3.1	1743	19 AAW98879	H. pylori GHPO 175
7	201.5	2.8	3248	17 AAR99795	Kinetochore protei
8	194.5	2.7	2482	16 AAR72826	Human mitotin. Ho
9	194.5	2.7	2482	19 AAW23996	Human mitotin amin
10	192	2.7	1703	21 AAG36714	Arabidopsis thalia
11	192	2.7	1710	21 AAG36713	Arabidopsis thalia

12	192	2.7	1744	21	AAG36712	Arabidopsis thalia
13	192	2.7	2954	20	AAY01632	Amino acid sequenc
14	187.5	2.6	1562	21	AAG39233	Arabidopsis thalia
15	184.5	2.6	1411	17	AAW02258	Nucleolar/endosoma
16	183.5	2.6	1521	21	AAG39235	Arabidopsis thalia
17	183.5	2.6	1528	21	AAG39234	Arabidopsis thalia
18	183	2.5	671	21	AAY99426	Human PRO1604 (UNQ
19	183	2.5	671	22	AAB66175	Protein of the inv
20	183	2.5	2485	21	AB18172	Plasmodium falci
21	182	2.5	1148	20	AAY07087	Renal cancer assoc
22	181.5	2.5	518	21	AAB53322	Human colon cancer
23	181.5	2.5	3685	10	AAP90290	Human Duchenne mus
24	181.5	2.5	3685	10	AAP90373	Sequence encoded b
25	181	2.5	1780	19	AAW53863	Human gravin poly
26	181	2.5	1780	21	AB15380	Human gravin prote
27	178.5	2.5	2368	18	AAW26663	Yeast checkpoint c
28	178.5	2.5	2368	20	AAW73893	Yeast MECL protein
29	177	2.5	2326	21	AAW71158	Rat phosphodiester
30	176	2.4	1319	14	AAW47043	Mammalian son of s
31	175.5	2.4	1091	16	AAW70982	Megakaryocyte stem
32	175.5	2.4	1091	16	AAW70983	Megakaryocyte stem
33	175.5	2.4	1429	20	AAW93941	Human brx protein.
34	175	2.4	669	19	AAW37483	Mouse liver cancer
35	174.5	2.4	1558	21	AB18324	Plasmodium falci
36	174	2.4	1336	16	AAW84638	msOS1 protein. Mu
37	174	2.4	1786	18	AAW24790	P. falciparum live
38	173.5	2.4	1257	21	AAW85281	C. albicans CADR32
39	172.5	2.4	1047	18	AAW01535	Sequence of the P1
40	172.5	2.4	1654	6	AAW50777	Cellular homologue
41	172.5	2.4	1865	17	AAW03515	Human DOCK180 prot
42	172.5	2.4	2096	21	AB441592	Human OREF ORF1356
43	171	2.4	718	21	AAG42569	Arabidopsis thalia
44	171	2.4	738	21	AAG42568	Arabidopsis thalia
45	171	2.4	763	21	AAG42567	Arabidopsis thalia

ALIGNMENTS

RESULT 1

ID AAY94702

AAAY94702 standard; Protein; 1391 AA.

AC AAY94702;

XX 01-DEC-2000 (first entry)

DT Human androgen shutoff gene 3 (AS3) protein sequence.

DE Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;

KW Chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;

KW diagnosis; treatment; cytostatic; human; ss.

XX Homo sapiens.

OS Homo sapiens.

Key Location/Qualifiers

FT Region 419..459

FT /label= "Subdomain I"

FT /note= "Hank's conserved region"

FT Region 419..425

FT /label= "Beta strand 1"

FT Region 426..433

FT /label= "Mg-ATP binding loop"

FT Region 453..459

FT /label= "Beta strand 2"

FT Region 472..482

FT /label= "Subdomain II"

FT /note= "Hank's conserved domain - beta strand 3"

FT Region 489..498

FT /label= "Subdomain III"

FT /note= "Hank's conserved region - alpha helix C"

FT Region 509..516

FT /label= "Subdomain IV"

3313 HGI-----DAMLLNIYHYFEOLFSLKSVSEAIHQKSLNSIKERILLMSWKDT-- 3360
408 AMGLAQIYKVKALQSAQKDAKJAWIKDKLLHYYONSIDDRLLVERIFAQYVPHN 467
3361 -----NVALKESAKSHAEF-----KVLHYRE-----VLQPVSSYLSQKHD 3400
468 LETERMKCLYLYAFDLNNAKALNEM--KQCNLLRHQV-----DLDLQIKPKRTDAS 521
3401 WD-----SLDTEENSA---MVAKKVNLSPSYIEKMDTEIMKLV--PVRFEN 3443
522 VKAIFSKVMVTRNLDPDK-----AQDPMKFTQVLEDE-----KIRKO 562
3444 TPTTLRLMTLFAVKEPGSTFTNMVSNLITDARELMKLTPTETINDNLSEIKHLSRKH 3503
563 LEVLVSPTCSQAECVREITKKGNGKQPTNPFFLEMIKFLIERAPVHIDTESALI 622
3504 L--LLTETFTKLAFGQIRV--KAGIEENLSN-----LRNLLAVIPTFPVTSLSIEKVD 3554
623 KOVKNKSID-----GTADDEDEGYPTDQAIKAGLELKLVSFTTHPIFSHSAE-----T 669
3555 RSLMKSLDFPKFOTLAGHQHNDLSP--EVQKGVGL-----FNSMLSIQLGERAQLVE 3606
670 FSLACLK--WDDEKVAEAAQIFKNTGSKTEEDPPHRSALLPVHLHKKSGPPRQAK 727
3607 FTNELALKNVYSEVGVNGSPLESEFNNS-----SPNEVSN--LGYDHDHFENRA--QAV 3655
728 YATHCHIAIFSSKETQFAIFEPLHKLSDPSNLEHLITPLVTIGHIALAPQFAAPWKS 787
3656 SMCQIYIAVIOK-----HSSISPT-----ASFQSIGH-----ELSRFA----- 3689
788 WVATFIVKDLMLNDRLPG-----KKTTLKLVVPDEEV 818
3690 -----DLLSNKLPSSIPYASADKYSSIRDQKGINDLIEYCRKRTTEL--PELSY 3739
819 SPETWYKIOAKMMVRLGKNNHKSSTSTLRLTTILHSDGDLTEOGKISKPDMSRL 878
3740 CFKHLVSLQSLKSIPTQVLDLND-----EFLNLMFNVLNFDLSLSIESTATKMMRTF 3793
879 R-LAAGSALVKLAQEPYHEITLEQYQCALAINDECYQVQVFAQKHLKGLSLRLP- 936
3794 KELAETSSEIENSS--CFSKVLRAFNLKFSQMKLSLSEKLSKSSVDKMSQCLLMFLFVP 3851
937 -----LEY-----MAICALCAKDPVKERRAHARCLVK 964
3852 CEQFINLAESVLDYFINVHNSNLSLSKISTLFFWYANNFGSPDLPOEGSKNSGE--- 3907
965 NINVRREYLKQRAV-----SEKLSLLPEYVYPTIHLHADDPVVKVQDIE--OLKDV 1017
3908 -----LESCTGLSGVGGAEDITWLN--DDDLEELANEEDTANGSDLDESEAREL 3956
1018 KECLWFLV-EILMAKNENSHAFIRKVENIKOTKDAQG---PDDA--KMEKLYTVCDV 1071
3957 ESDMGVTKDSVVSENEENSDS-----BEENODLDEEVNDIPEDLSNLEKLV----- 4004
1072 AMNIIMSKTYSLESFKDVPFLPFPFTQDKNFSTKNYLPPEKMSFFETPGKPTTNVL 1131
4005 -----DEP-----NEEDLLETEQS-----NEQ 4022
1132 GAVNKPLSAGKQSQTKS-----SRMTVGNASSSSPPSGRIKGL--DSSEMDHSHENED 1186
4023 SAANNESDLVSKREDDNKALDQKEDEEEMSDVGIDDETDQIDQENNSOPPPENED 4082
1187 YTMSSP---LPGKSKDRDDSDLVSELEKPRGRKKTPTVEEKLGMDDLTKLVOEQK 1242
4083 H-LJLDPEDLKLDEKGDVSKDSLEDDMDMEADENKEEADAERKDEP--MODFEDPLEENN 4139
1243 PKGSQRSRKRGHTASDESQOMPEERKLKEDILENEDEQNSPPKKGKRGPRPKPLGGGTP 1302
4140 TLDEDIQDDFSLADEDEKM--NEDGFENYQENESTEDGVKSDDELEQ-----GEVP 4192
1303 KEPTWTKSKKSKGPPAPPEEEEEERQSGNTBQKSKQHOHVRVRAQRAESPSS 1362
4193 EDQAIIDHPKMDAKSTFASAEADEENTDKGIVGENEELGEDCAESGVGRNGTADGEFS 4252

QY 1363 ATESTQSTPQKGRGRPSKTP 1382
DB 4253 SAEQVQ-----KGEDTSTP 4266

RESULT 15

T18296

myosin heavy chain - Entamoeba histolytica

C:Species: Entamoeba histolytica

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000

C:Accession: T18296

R:Guillen, N.

submitted to the EMBL Data Library, February 1997

A:Reference number: Z18865

A:Accession: T18296

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2139 <GUI>

A:Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1

C:Genetics:

A:Gene: mhca

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 3.1%; Score 223; DB 2; Length 2139;
Best Local Similarity 19.6%; Pred. No. 0.0086;
Matches 292; Conservative 233; Mismatches 570; Indels 396; Gaps 62;

QY 2 AHSKRTNDGKI-TYPPGVKEISDKISKSEEMVRR-LKVVVKTMDMDQSEEEKELYNL 59
DB 894 ATAKTELEAKTDLEDKISELSAELDKQELNLAIE---NLEEDKEELKETIDNL 950
QY 60 ALHASDFFLKHPGKDVRLVLAACLADIFRIYAP-----EAPYTPDKLKDIFMT 111
DB 951 KGLDKDS---KLKGEDLEVEITELNSQINTLNATVDKDKTIAEMQESIDEKEDETKLK 1007
QY 112 ROLKGLGDTKSPQFNRYFYLLENIAWV-----KSNICEFEEDSNEIFTQLYRTLF 162
DB 1008 GDIKLLEEKEDD---LEQDRADVSA TKDDIAKLNKITECEDAKDEIAKLEOELE 1060
QY 163 SVINNGHN---QKVHMHVMDLMSIIC-----EGDTVSQELLDLTVLVNLVPAH 207
DB 1061 DEENKKNKDLTNELOQTQLKGETEKSLAAQVAATKASDERDTLSQNLNLEKLT--- 1115
QY 208 KNLNKOAYDLAKALKARTQAIEPYITTFNQVLMGLKT7SISDLSEHVFDLILELYNIDS 267
DB 1116 KNLTKTKADLEK-----KISGLKQDYEDL-----EDDK 1143
QY 268 HLLLSVLPQLEPKLSKSNDEERLOVVKLLAKMFGAKDSELSQNKPLWCYGLGRFNDIHV 327
DB 1144 NKIEGLRNAQRKIKELDDE---ITK-----GADVSYLQKQKEEYESQIAKM----- 1188
QY 328 PIRLECVKFAHCLMNHPLDAKDLTEYLVKVRSHDPEAIRHVDIVSIVTAARKDILLVND 387
DB 1189 -----QEEKEATGNVKNKEKTIKEKELEIQS- 1215
QY 388 HLLNFVRETLTKRWVRKRAMGLAQIYKKY-ALQSAAGKDAKQIAIMTKLLHIYYQ 446
DB 1216 -----LQEKLDTEVEKEDEAKKKKEIEKMKALQEEKENVESSKNSTEKDK----- 1262
QY 447 NSIDDRLLVERIFAQYVPHNLETTERMCLYYLYATLDLNAVKALENNKQNLRLHOV 506
DB 1263 KKLED-----NLKDTQ-----KKLDDMTADNEKLKAKA 1290
QY 507 KOLLDLIKOPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDEKIRKOLEVL 566
DB 1291 KDLEAQLNEVQ--DNHEKAV-----ADAEKLNKKAQSKDELNSLKAELEAL 1335
QY 567 VSPTCCKQAECQCVREITKKGNGKQPTNPFFLEMIKFLIERIAPVHIDTESIALIKQVN 626
DB 1336 T-----KAKSVESKNDSENEKAALS---BEIDQANEKLNKIQADLRKATADLQEAN 1385

623 LGVVTAPTSTNKNLSTDMFELFWHLITPIYEDVYFKYDSIHLQNLVFTVLHLLIGKNA 682
409 MGLAQIYKKY-----ALQSAAGKDAK---QIAWIK-----DKLLHIYQNSIDRLLY 455
683 DVALERKYKKHHPMSVITASBGVLKDISSLPPOI--IKREYDKIMVFEQ-----V 733
456 ERIFAQYVMPINLETTERMKCLYLYATLDLNAVAKALNEMKQCQNLRLHQVKDLDLLIKO 515
734 EVAISNVNLADHLITS-----LKHLPED----- 757
516 PKTDASVKAIFSKVMTIRNLDPGKQADFMKKFTOVL-----EDEKIRKOLE 564
758 RKDQTHLESFSLILKYTONKNKDPIDPFCAGVTSFVYTFDLFLRKNDSLVNFNIQ 817
565 VLVSPSCCKOAEQCVREITKKLGNKQPTNPFLFMKFL-----ERIA 609
818 ISKVGISQGNMTDLLKDVIRKARN---ETSEFLIEKFELEDDDKTEVYAQNVMVGSTLL 874
610 PVHIDTESIALIKOVKNSIDGTADDEGYPTDQAIRAGLELLKVLSTFTPIFHSNET 669
875 PPNISFREFOQLANIVNK-----VPNENSIENFLDLCLKLSF--PVNL----- 915
670 FESLIACLKMDKEVAEAAALQIFKNTGSKIEEDF-----PHIRSALLPVLHHS 718
916 FTLLHVSWMNNFYFIQSYVSKNE-NKLAVDLITLLKTSPLGNPELFSGLLPFL----- 970
719 KKGPPROAKY---AIHCIIHA-----IFSKKE- 741
971 -----RRNKFMIDILEYCIHSPNLLNSIPDLNSDLLKLLPRSRASYFAANIKLFKCEQ 1025
742 -----TOFAIIEPLHKSIDPSNLE--HLITPLVTIGHIALLAPODFA 782
1026 LTLVRWLKGOOLEQNLQNFSEIENVLQNASD-SELEKSEIIRELL---HLAMANPIE-- 1079
783 APWKSQWATFIVKOLLMN--DRLPGKTKTKLWVPDEEVSPETWVKIOAKMVRWLLGMK 840
1080 -PLFSGLLNFCKKNMADHLDPCGNTSEVLF---KISPELLIKLLTYK----- 1125
841 NNHKSQSTLRLTLTLHSDG-----LTEQGIKSKPDMSRLRLAAGSAIVKLAQEPYH 896
1126 --EKPNKLLAAVIEKIEGNGDDVILELEKIIIOKE-----IOILEKLKEP--- 1170
897 EIITLEQYOLCALAINDECYO-----VROVFAQKLHGLSLRLPLEYMAICALCAKDP- 950
1171 ---LLVFFLPVPSNMOKHKSNMURELVLLYTLPLRSAAKFFSMILISILPPNP 1226
951 -----VKERRAHAROCQVKNINVRREYLKQ-----HAAVSEKLLSL 986
1227 YQIDMVNLLDLIKSHNRKFKDRTY--NATLKTIG--KWIQESGVVHOGDSKEIEA 1281
987 LPE-----YV-----VPYTIHLAHDPPYKVDIEQLKDVKEC 1020
1282 IPDTKSMYIPCEGSENKLSLQKRVDSODIQVPAT-OGMKPEPPSSIQSISAKD-SDS 1339
1021 LWFVLEILMAKNENNSHA-----FIRKVENIKOTK-----D 1052
1340 ISLKNTAIMNSQOESHANRSRIDETLEEVNDESIREIDQOMKSTQOLDKRVANHSNIC 1399
1053 AQGPDDAKMNEKLYTCVDMANNIMSKSTTYSLESKDPVLPARFFPTQDKNFNNTKNYL 1112
1400 STKSDEVDVTE-LHESIDTQSSEVNAVQPIEVLTSELKAVTNRSIKTNPDPHNVVNSDNL 1458
1113 PPEKSFPTPKPKPT-----TNVLGAVNKPLSSA----- 1141
1459 KRPSKETPTSENKRKSGHETWMDVLVSEQAVSPSSDVICTNIKSIANEESLALRSIK 1518
1142 -----GKQOTKSSRMETVSNASSSSNPSSPGRIKGLDSE----- 1178
1519 VETNCNENSLNLTLDQOITTKEDGKGQVHVRQENQESMNNKINSKSFQODNIAQVKS 1578
1179 -----MDHSENEIDYMSPLPGKSKDRDSDLVRSR-----LEKPRGRKKTPTVTEOE 1227
1579 VKARPNNEGENNDYACV-----EQASPVREVPDGIQIPSGTILLNSSKQTE 1628

QY 1228 KLGMDDTK-----LVQEOKPKGSQRKRGRHTASDEEQMPPEKRLKEDILEN--- 1277
D 1229 KSKYDDLRSDEDEHGTVAQEKHQVGAINSRNKNNDMDSTPIQGTTEESREVVMTEGIN 1688
QY 1278 --EDEQNSPPKPKGRPPKPLGG-----TPKEPTMTKSKGSKKSG 1320
D 1689 VRLDESGTCELNKLKG-----PLAGDKDANINDDFVPVEENVROEGFLKSEHAVSKETG 1744
QY 1321 PPAPEEEEEE-----ROSGTEQKSKQHRVSRRAQOAEESPSS 1362
D 1745 -----LEEQPEVADISVLPEIRIPFNLSLKMQGSQSKIEKLRQLORNELMPDPPRM 1799
QY 1363 AESTQSTPQKGRGRPSKT 1381
D 1800 T-ENTNINAQNGLDTPKPT 1817

RESULT 14

T41581

hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000

C:Accession: T41581

R:Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z22002

A:Accession: T41581

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4717 <MUR>

A:Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08

A:Experimental source: strain 972h(-)

C:Genetics:

A:Gene: SPDB:SPCC737.08

A:Map position: 3

C:Superfamily: Saccharomyces probable membrane protein YLR106c

Query Match 3.1%; Score 223.5; DB 2; Length 4717;
Best Local Similarity 18.4%; Pred. No. 0.023;
Matches 290; Conservative 249; Mismatches 572; Indels 469; Gaps 71;

QY 24 DKISKEEMVRRLKVVVTFMDMDQDSEEEKELYLNLAHLASDFLKHGPKGKDVRLVACC 83
D 2935 DVDS--NLIEQIKV-----LDLNDISIRFEIYFLASRLCSE---KQHSDDTH-----S 2977
QY 84 LADIFRIYAPEAPYTPDKLKDIFMTITROLKGLDTSKQFNRYFYLLLENIAWVK---S 140
D 2978 LANSEVLLANEF-YIHNAKIKQ-----KELEEEKNRLYRQREFNFDKN-DYLKVFIN 3029
QY 141 YNICFELEDSNEI-----FTQLYRTLFVINGNHQKVVHMHVMDLMSI-----I 185
D 3030 YDDEVEPEVEPEVIERKRFLOQAFWLSLYNEYSEKMNVPLEQLMNTGSLYAKKIKV 3089
QY 186 CEGTVTSQELDTVLVNLVPAHKNLKOAY-----DLAKALLKR 224
D 3090 KNPDMIASSGPDIVSVVLMGMVKSTNERQYWTTPVYFNYPNPSPKATEVRDLKIVESR 3149
QY 225 TQAIEPIVITTFQVOVLMLGKTSISDLS-----EHVFDLILEL-----YNI 265
D 3150 AISLIKWPENFVIRGLKDAIDAILNSPFPSPPIAEYLSKLERVHLLSEWEKLASREYSL 3209
QY 266 DSHLL---SVLPQLEKLSNDNEERLQVYKLLAKMFGA-----KDSLEA 308
D 3210 ANEMDLIKKIIDNRKPFELSNWNLKLEEVKLSERVYPRLYSILOFIIKLPFFENSKFT 3269
QY 309 SQNPLWQCYLGRF-NDIHVPIRECVKFASHCLMNHDPDLAKDLTEYLKVRSHDPEAIR 367
D 3270 KQNCESASIIQVITDITLVGEFQCLK-----CLLSFSQHAASL-----RIC 3312
QY 368 HDVIVSVITAAKKDILLVN-----DHLLNFVRERTLDKRWVRKE 407


```
QY 720 KQPPQAKYAIHAIHFSSKQTFQAFIPELHKSLOPSN-----LEHLITPLVTIGHIA 774
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
931 KDDLKQLESQI-----ERQKSDIHDVTVMNIDTQQLRNALESKQHOETINTL- 983
QY 775 LIAPDOFAAPKSWATFVIFKLLMNDRLPGKTKLWVPDEVSPTVVKVQAKMMVYR 834
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
984 -----KSKISEEYSRLHM-----BENTGETKDEFQ-----Q 1010
QY 835 WLGMKNHNSKSTSLRLTLILHSDGLTBOGKISK--PMSRLRLAAGSAIVKLAQ- 891
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1011 KWGIDDKODLEAKNF-QLTADVNDNEIEQORKIFSLIQEKNELOQMLESVIAKEQL 1069
QY 892 -----EPCYHEITLBOYLQALAINDECYQVQVPAQKLHKLGLSLRLPLLEYMAICALC 946
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1070 KTDLKENIEMTIENQELRL-----LGDELKQOEIVAQEKNAHK--EGELSRCTCDRL 1122
QY 947 A--KDPVKERRAHARQCLVKNINVRREYLUKHAAYSE-----KLSLLPEYVVPYIT 996
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1123 AEVEEKLKEKSQLQKQOOLLNVQEMSEMOKKINEIENKLNKELTLEHMETERL 1182
QY 997 HLLAH-DPQVVKVODI-EOLKDVKECLWFVLEILMAKNENNSHAFTRKMVENIKOTKDAQ 1054
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1183 ELAQLNENYEVKSTKERRKVLKE-----LQKSPETERDHLRGYIRETEATGLQTKBEL 1237
QY 1055 GPDDAKMNEKLYTVCDAVAMNIIMSKSTTYSLESPPKDPVLPARFFTPQDKNFSNTKNYLP 1114
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1238 KIAHILHKEHQETIDELRSV--SEKT-----AQIINTQDLEKSHTK--LQE 1280
QY 1115 EMKSFTPGKTKTNNVLGAVNPLSAGKQSTKS--SRMETVSNASSNSPSSPGRIKG 1172
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1281 ETP-----VHHEQEQLLPNVKVSQETQTMNELELLTEOSTTKDSTTLARI-- 1326
QY 1173 RLDSSEMDHSE-NEDYTMSSPLPGKSKDKRDSLDLRSLEKPRGKKTPTVE----- 1224
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1327 -----EMERLNEKEFQESQEKSLTKRDNKLTIKAEVKKHDKLKEHIRETIAKIQE 1381
QY 1225 -----QEEKLGK-----DGLTKLVQEQ---KPK-----GSRKSRKRGH-----TA 1256
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1382 SOSKQBSLNMKEKONETTKIVSEMPQPKDPSALLRIEIMGLSKRLQESHDEMKSYA 1441
QY 1257 SPSDEQ-----WPEKRLKEDI-----LENEDEQNSPKKGRGPP--KPLGGGTP 1302
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1442 KEKDDLQRLQEVQESDQLENKIRKIBAKHLETEBELVAHCCLEQEBETINELRVNLS 1501
QY 1303 KEETPMKTSK-----GSKKKSGPPAPEEEEE---EROSGNTQKSKOHRVSPRAQ 1354
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1502 EKETEISTQKOLEAINDKLQIBYKEBQNLKQISEVOENVNELQKPKHRKAKD 1561
QY 1355 RA-ESPSSAIBST 1367
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1562 SALQSTESKMLEIT 1575

RESULT 13
S46157
R1F1 protein - yeast (Saccharomyces cerevisiae)
N1Alternate names: protein YBR1743; protein YBR275c; RAp1-interacting factor
C1Species: Saccharomyces cerevisiae
C1Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
C1Accession: S46157; S46156; S44537; S25704; S39131
R1Brandt, T.; Christiansen, C.; Holmstrom, K.; Kalliesoe, T.
submitted to the Protein Sequence Database, August 1994
A1Reference number: S46157
A1Accession: S46157
A1Molecule type: DNA
A1Residues: 1-1096 <BRA>
A1Cross-references: EMBL:Z36144; MIPS:YBR275c
A1Experimental source: strain S288C
R1Aigle, M.; Baclet, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.
submitted to the Protein Sequence Database, August 1994
A1Reference number: S45940
A1Accession: S46156
A1Molecule type: DNA
```

```
A:Residues: 1005-1916 <AIG>
A:Cross-references: EMBL:Z36144; MIPS:YBR275c
A:Experimental source: strain S288C
R:Holmstrom, K.; Brandt, T.; Kalliesoe, T.
yeast 10(Suppl.A), S47-S62, 1994
A:Title: The sequence of a 32420 bp segment located on the right arm of chromosome II
A:Reference number: S44537; MUID:94378722
A:Accession: S44537
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-1096 <HOL>
A:Cross-references: EMBL:X76053; NID:g600025; PID:g429120
A:Experimental source: strain S288C
R:Hardy, C.F.J.; Sussel, L.; Shore, D.
Genes Dev. 6, 801-814, 1992
A:Title: A RAP1-interacting protein involved in transcriptional silencing and telomer
A:Reference number: S25704; MUID:92249772
A:Accession: S25704
A:Molecule type: DNA
A:Residues: 1-579,'S',581-731,'A',733-1916 <HAR>
A:Cross-references: EMBL:X66501; NID:g4327; PIDN:CAAA7121.1; PID:g4328
C:Genetics:
A:Gene: SGD:RIF1
A:Cross-references: SGD:S0000479; MIPS:YBR275c
A:Map position: 2R
C:Superfamily: Saccharomyces cerevisiae RIF1 protein
C:Keywords: transmembrane protein
F:263-279/Domain: transmembrane #status predicted <TM1>
F:296-312/Domain: transmembrane #status predicted <TM2>
F:438-454/Domain: transmembrane #status predicted <TM3>
F:663-679/Domain: transmembrane #status predicted <TM4>
F:788-804/Domain: transmembrane #status predicted <TM5>
F:905-921/Domain: transmembrane #status predicted <TM6>

Query Match 3.2%; Score 227; DB 2; Length 1916;
Best Local Similarity 17.4%; Pred. No. 0.0048;
Matches 316; Conservative 267; Mismatches 612; Indels 624; Gaps 80;

QY 3 HSKTRTFDQKIY-----PPGVKEISDKI-----SKEEMVRRLKMW--- 38
DB 183 YKTRVNPRALEYVSGETHGLVDNESVEFKEIEGGLIRLQESDYVARRFEVYATF 242
QY 39 -----VKTMDMDQDSEEEKELYLNLA-----LHLASDFFLKHGPKDVLVACCL 84
DB 243 NNIIPILTTKNVNDQKFNILIVNIESIETICIPHQLAQDTLLSSSEKKNPFVI--- 298
QY 85 ADIFRIYAPEAPYTS-----PD---KLKDF-----MFTR 112
DB 299 ----RLYVQIVRFFSAIMSNEFKIVKWLTKRPDLVNLKVIYRWTTGALRNSNKIIITA 354
QY 113 QLKGLDTPKSPQNFYFYLENI-AWVKSYNTCFELEDSENEIFTQLYRTLFSVINGHNO 171
DB 355 QVSFLRDEK---FGTFFLSNEEIKPIITFTTEINSHNLIYEKL-----LLIRGFLS 405
QY 172 KVHMHMVDLMSS-----IICEGDTYSQELDLVNLVPAHKNLKNQAYDLAKLL 222
DB 406 KYPKLMIEFTVSWLPGEVLPRIIGDEIYSMKILITSIVLL-----ELL 450
QY 223 KRTAQAIPEYITTFNQVLMGKTSISDLSERFVDLILELY---NID-----SHLLSLVLP 275
DB 451 KKCLDFVDEHERIY--QCIMLSPVCETIPEKFLSKPLANSYDSANLDKVTIGHLLTQQIK 508
QY 276 QLEFLKLSNDNEERLQV-VKLLAKMF--GAKDSELASQNKPLW-----OCYLGRFNDIHV 327
DB 509 --NYIVKNDKNIAMDLWSMTGLLYDSGRVYDLTSESKNVWFDLNNLCFTINN----HP 562
QY 328 PIRLECVK--FASCL-----MNHPLDAKLDTLYLK-----VRSHDPEAIRHDV 370
DB 563 KYRLMSIKVWRITTYICTKISKQKQEGNKSLSLLRTFQMTPLPYVNDPFSAREGIYHL 622
QY 371 IVSIVTA--AKKDIL-----LVNDHLLNFVRERTLDKRWVR-----KEA 408
A:Molecule type: DNA
```


943 LSDKYKKAEPFAMVVKVLKEQNSERIASLESELNSLAVKVERKCESEKLSGEVEHLKQ 1002
QY 629 IDGTADDEGV-----PTDQAIRA--GLELLKVLFSFTHPIFSHSAET 669
DB 1003 LDDSSQRHKEALAKNIEIKOLINAKKATSDLAIKSEMGAOLQKAVD-THKSEFALON 1061
QY 670 PES-----LLA-----CLKMDKVAEAL----- 689
DB 1062 ELSRSLDLALKEGEVERLNKEAALROBEIOOOOQTITKLEETEALAAALDKVALOEKE 1121
QY 690 ---OIFKNTGSKIEDPHIRSAALLPVLLHHSKK-----GPPRAKVAIHCHIAIFSSK- 740
DB 1122 IKQVQATKGA--EKEMAKLAS--VISEKSRIBCLEQIQKRDLSICIQHQHOSKL 1175
QY 741 -ETQFAQ-----IFEPLHK-----SLDPSNLEHLITPLVTIG 771
DB 1176 GESQGLQALJADLEKKCKEOKELICEAQNAKAAEAKTLASEKASVSERQLEGIOALEIEIG 1235
QY 772 HIALAPD---QFAAPW-----KSWATFIVKDLIMN-DRLPGRKTT-----KLWP 814
DB 1236 KERQACDLQKLESLMAVOBEKETELQALKKELFKVQOELEQSQTSFTDSSGGEALLYLS 1295
QY 815 DEEVSPETMVKIOAKMVMRWLLGKMKNNHSGSTSLRLTLTILHSDGD-----LT 865
DB 1296 EAQERQALTAKEQAEOYKEIEHMKNEVNSLOAEIKILSSKVTTNNEVSVDFEQRLIK 1355
QY 866 EQGKISPD---MSRLRLAAGSAIVKLAQEPYHEIITLEQYOLCALAINDECYQVQVFA 923
DB 1356 ETSKSALEKMKWLHWELEASFRELLEKNCAIDCLITEAQNLAGEA--DOORMAVDSLQ 1413
QY 924 QKL-HKGLSRRLPLEYMAICALCAKDPKERRAHARQCLVKNINVRREYLUKQHAASEK 982
DB 1414 QKLSKAEHTNHTLOQEIQAOMKNAE---KEQQICSLQ---QNLKSNQSLLEEFASLKH 1467
QY 983 LLSLPEVVVPTTH---LLAHPDYPVKVQ--DIEOLK-DVKECLWFVLEILMAKNENNSH 1037
DB 1468 YOEIIAERDLMOEKHQBELSHKKLTERFOAEKAKEDMTEIV--LLKEKLHNOELQLH 1525
QY 1038 AFIRKVENIKQTKDAQDPDAKM--NEKLYTVCDAVAMNIIMSKTTSYLESKPDVPLA 1095
DB 1526 KFQSENSVSLTOISHLQOVNSOLLGANSLSOISDOGAKKLESEMSTLK-BQHKE----- 1579
QY 1096 RFTQPKNFNTKNYLPPEMKSFETGCKPKTNTVGLAVGNKPLSAGKQSKTSRSMETV 1155
DB 1580 -----EMKTLRL-----OYEKTLREGNKQVQETSLOQETV 1609
QY 1156 SNASSSSNPPGRIKGRDSEMDHSENEYDTWSSPLPGKKDKRDDSDLVRSLEKPR 1215
DB 1610 -----TSKYDH-----VSKSVLKQD 1624
QY 1216 GRKKTPTVTEQEK-----LGMDDLTK-LVQEQPKGQSRSR---KRGHTASESDQOMPEE 1267
DB 1625 -----KTFQBEKRLLLQVQLNKLQSQBEKTRISQOQKLQREGTHEEADK-----SH 1674
QY 1268 KRLKEDILEDEQNSPPKGRPRPKPLGGTTPKEPTMKTSKKSKKSGPPAPEEE 1327
DB 1675 KRVLK--LESQLEQOOTQAVEHYKAQMEK-----AKVHYDAKKON-----QE 1714
QY 1328 EEEERQSGNTEQSKSKOHRVSRRAQORAESPE 1360
DB 1715 LSELQSHIKQOEHLSEN-----ADLKAESQ 1742

RESULT 12

S28261

centromere protein E - human

N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E

C:Species: Homo sapiens (man)

C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001

C:Accession: S28261

R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.

Nature 359, 536-539, 1992

A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.

A:Reference number: S28261; MUID:93024922
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CEPPE
A:Cross-references: GDB:361164; OMIM:117143 -
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: Atp; coiled coil; microtubule binding; mitosis; nucleotide binding; p-loop
F:7-335/Domain: kinesin motor domain homology <KMT>
F:86-93/Region: nucleotide-binding motif A (p-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: Atp (lys) #status predicted

Query Match 3.28; Score 229; DB 1; Length 2663;
Best Local Similarity 18.2%; Pred. No. 0.0059;
Matches 275; Conservative 268; Mismatches 537; Indels 434; Gaps 64;
QY 17 PGVKEISDKISKEEMVRRLKMWVKTFMDMDODSEE-----EKELYLNLAHLASD 66
DB 333 PYNVEVS---TDEALLKRYR---KEIMDLKKOLEVSLETRAQAMEKD---QLAQLLEEK 383
QY 67 FFLHPGKDVRLLVACCLADIFRIYAPEPYTSPDKLKDIF-MFITRLQKGL-EDTKSPQ 124
DB 384 DLLQK-----VQNEKIENLTRLMLVTSSSLTLQOELKAKR 417
QY 125 FNRFFYLENTAWKSYNICEFEEDSNEIFTQLYRTFLFSVINNGHNGKVVHMHVMDLMSI 184
DB 418 KRRVTWCGLGINKMKNSN-----YADOFNIPTN-ITTKTHKLSINLLREI 461
QY 185 ---ICEGDTVSQELLDVFL-VNLVPAHKNLKNQAVDLAKALLKRTAQAIPIYITTFNQV 240
DB 462 DESVCSSEDFSNLTDLTSLSEIENWPATKLLNQENIE----- 497
QY 241 LMLGKTSISDLSERHFDLILELYNIDSHLLSVLPQLFEFLKLSNDNEERLOVQVLLAKMF 300
DB 498 -----SELNSLRADYDNLVL-----DVEQLRTEKEEMELKLEKNDLDEFEALERKTK-- 545
QY 301 GAKDSELASQNKPLWQCYLGRFDIHVPIRLECYKFASHCLMNHDPDLAKDLTEYLKVRSH 360
DB 546 --KQEMQ-----LIHEISNLKLVKRVETNQ 571
QY 361 DPEAIRHDVITSVTAACKDILLVNDHLLNFVRERTLDKRWVRKKAAMGLAQYKKYA 420
DB 572 DLENEL-----SSKVLELREKEDQIKKLOFYIDSQK---LENIKMDLS-----YS 613
QY 421 LQSAAGKDAKQIAWIKDKLHIYYQNSIDDRLLVERIFAQYVPHNLETTERMKCLYYL 480
DB 614 LESTEDPKMQKQTLFDAETV-----ALDAKRESAFLRSENLELKERMKELATT 661
QY 481 YATLDLNAVKNALNEMWKCQNLLRHQVKDLDLILKOPKTDASVKAIFSVMVITR----- 534
DB 662 YKOME-----NDIQYQSOLEAKRMQVDLEKE-----LQSAFNEITKLTLSDIGKV 708
QY 535 -----NLPDPGKAQDFMKKFTQVLEDDKIRKOLEVLVSPTCSCKQAEGCVREITKLG 588
DB 709 PKDLLCNLEEGKITDLOKELNKEVEENEALREEV-ILLS---ELKSLPSEVERLKEI- 763
QY 589 NPKOPTNPFLEIMIKFLLERIAPVHIDTES-ISALIKOVNKSIDGTADDEGVPTDQAI- 646
DB 764 ---QDKSEELHIITSEKDKLFSVVVHKESRVQGLEEIGTKDKDLATTQSNKYKSTOQBOFQ 820
QY 647 -----RAGLELLKVLSTFTHPSFHSATFESLACLKMD----- 680
DB 821 NFKTLHMDFEQKYKMWLEENRMNQEIUNLSK-----EAQFDSLSLGAALKTSLSYKT 872
QY 681 ---DEKVAEALQIFKNTGSKIEDFPHIRSALLPVLHHS-----K 719
DB 873 QELOKRETVQERL--NEMEQLEQLENRDSPLQTVREKTLITEKLOOTLEEVKTLTQE 930

Db 47 VEODLLSSVOKALHPMRAVSAD--LLRNPDSVVRVSVVSCUTEIMRITAPAPY-NDQ 104
Qy 103 LKIDPMTITRQKLGLEDTKSPQNFYFYLENIWVSKSNICFELEDSNEIFTQYRILF 162
Db 105 MKDIFQVITAEPEKLADASSRSYRAEVLFTVAKVRSSVLMDLE--CDLVLEMFQRL 163
Qy 163 SVINNGHOKVHHMVLDLSSIIICGDTVSOELLDVNLVPAKLNKQAYDLAKALL 222
Db 164 KIIRDPHQLVLSMETIMITVIDSEVPMDLLELLTTVRKDSQDVSPAALTIVERVL 223
Qy 223 KRTAQAEYITTFNFQVLMGLKTSISDLSEHFVLDILELYNIDSHLLSLVLPQLEF--- 279
Db 224 SSTCKLQPCIM---EALKSSGTS-----LDMS-----PVVSSICQSEFAT 263
Qy 280 -----KLKSNONEERL---QVY--KLLAKMGAKDSELASQNKPLWQCYLGRFNDIHVP 328
Db 264 QAHNDVKPKDNEADEKISGQVVPDSLEDKLNLGLSRKTRSKRSARGGTRRANGD--- 320
Qy 329 IRLECYKFASHCLMHPDLAKDLTEYLKVRSHDPPEAIRHVDIVSVITAAKDDILLVNDH 388
Db 321 ---EKVITANEGLESTD--AETASGTRKRGWKPKSLMNPBEGYFKTSSSKV----- 370
Qy 389 LLNFVRETLDRWRVRKEMAGLQIYKKYALQSAAGKDAQAQIAWIKDKLLHIYQNS 448
Db 371 -----QEKELG-----DSSLGKVAKKVP----- 389
Qy 449 IDORLLVERIFAQYVWPHNLETTERMKCLYLYATLDLNAVNALNMWKCQNLLRHQVKD 508
Db 390 -----LPSKVGQTNQ-----SWISLSSGSGRARTGSRKRSR- 420
Qy 509 LLDLIKQPKTDASVKAIFSKVWVITRNLPDQKADFMKFTQVLEDDK--IRKOLEVL 566
Db 421 ----TKMETDHDVSSVATQ-----PAKKQT-VKKTNPADKDLTKSNVKKHEDGI 465
Qy 567 VSPTECKQA--EGCVREITKKLGNPKQTPNPLE--MIKFLERITAPVHIDETESALI 622
Db 466 KTGKSKKEKADNGLAKTSKAK-----PLAETMMVKPSGKKL--VHSDAK----- 508
Qy 623 KOVKNKIDGTADDEGVPTDQAIRAGLELLKVLSTPHSFHSATFESLLACUKMDE 682
Db 509 ---KKNSGASMD---TPIPOSSKS-----KKKDS 532
Qy 683 KYAEALQIFKNTGKTEEDFFHRSALLPVLHHSKKGPPQAKYAIHCHIAIFSSKET 742
Db 533 RATTPATK-----KSEQAPKSHPKMKR-----IAGEEV 560
Qy 743 QFAQIPEPLHKSLDPSNLEHLITPLVTIGHIALPADQFAAPKSWVATFIVKDLLMND 802
Db 561 E-----SNTNEL-----GEE 570
Qy 803 LPQKTKTKLWVPDEVSPTMVKIQAIKMWVRWLLGMKNHNSKSGTSLRLTLTLHSDG 862
Db 571 LVGKRVNWMPLDKKFEYGVKSYCRVKKN-----HQ-----VTYSOG 608
Qy 863 DLTEQCKISKPDMSRLRLAAGSAIVKLAQPCYHEITLEQYQCALAINDECYQVRQVF 922
Db 609 DV--EELNLKK---ERKFIIEDSSASEDDLESTPLSAF----- 646
Qy 923 AQKLHKLRLPLPLEYMAITCALCAKDPVKERRAHARQCLVKNINVRREYLKQHAASEK 982
Db 647 -----VPE-----SSSEPVRSMSQMTKKKD-----IQREKSKRKIVSKN 661
Qy 983 LLSLLPEYVVPYTHLLAHDPPVVKVQDTEQKQVKECLWFLVLEILMAKNENNSHAFIRK 1042
Db 662 -----VEP-----SSSEPVRSMSQMTKKKD-----S 682
Qy 1043 MVENIKOTDAQDAPDDAKKNKLYTVCDVAMNIMSSTYSLESKDPVLPARFFTPQD 1102
Db 683 VTDISIKQTKRTKALKAVSNP-----ESTGKNLKSUK-----KLNGEPD 723
Qy 1103 KNFSNTKNVLPPEMKSGFFPGPKTTNVLGAVNKLSSAGKQSKTSRMETVSNASSSS 1162

Db 724 KTRGRT-----GKKQ-----KVTOAMHRKIEK----- 745
Qy 1163 NPSSGRIKGRILDSSEMDHSENEYTMSSPLPGKKSKDRDDSDLVSRSELEKPRGRKKTVP 1222
Db 746 -----DCDEQEDLETKEDEDSUKL--GKESDA--EPDRMEDHOELPENHNVTETK 790
Qy 1223 TPOEEKLGMDDLTKLVQEQPKGSQSRKRGHTASBESDQWPEEKRLKLEIDLE--NEDE 1280
Db 791 TDGEE-----QEAAKEPTAESKTNGEENAPETDCKEHSKLEPNAEPKSDGE 839
Qy 1281 QNSPPKK-----GKGRPPKPLGGGTPKKEPTMKTS-----KKGSKKSGPPAPEEE 1328
Db 840 EQEAAKEPNAELKTDGQENQAAKELTAERKTDDEEHKVADEVEQKSKETNVEPEAGEE 899
Qy 1329 EEROSGNTQOKSKQHRVSRRAOQRAESPRESSAIESTQSTPQKGRGPRSKTPS 1383
Db 900 QKSVEEPNAEPKTK-----VEEKESAKEQTADTKLIEKEDMSKTKGEIDKETYS 949
RESULT 9
S38173
myosin-like protein MLP1 - yeast (Saccharomycetes cerevisiae)
N:Alternate names: protein YKR095w; protein YKR415
C:Species: Saccharomycetes cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale
submitted to the Protein Sequence database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR095w
R:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomycetes cerevisiae chr
A:Reference number: S40644; MUID:94205265
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
R:Experimental source: strain S288C
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A:Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549
A:Accession: S31207
A:Molecule type: DNA
A:Residues: 1-300, 'A', 302-1875 <KOE>
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C:Genetics:
A:Gene: SGD:MLP1
A:Cross-references: SGD:S0001803; MIPS:YKR095w
A:Map position: 11R
Query Match 3.3%; Score 235.5; DB 2; Length 1875;
Best Local Similarity 19.5%; Pred. No. 0.0018;
Matches 301; Conservative 246; Mismatches 602; Indels 391; Gaps 72;
Qy 19 VKEISDKISKE-----EMVRLKVVVKTMDMDQDSEEEKEL-VLNLAHLASDFLKH 71
Db 278 IGLSDLSNSEQEFAEMSLKQRLVDLLESQNAVKEELNSRELTAKVIAADSSKKQT 337
Qy 72 PG-----KDVRL--VACCLADIFRI-----YAPEPTYSPDKLKIDFMFITRLKGLLE 118
Db 338 PENEDLLKELQTLKEKLAQCEKLECLSSITDEADENLSAKSSSDFIFLKQLIKER 397
Qy 119 DTK---SPQFNRYFYLLN-IANVSYNTCFELEDSENEIFTQYRFLFSVINNGHOKVH 174
Db 398 RTKEHLQNIQIETFIVELEHKVPIINSFK---ERTDMLNELNNAALLLETSTSEKNAKVK 454

[illegible]

```

Query Match      4.3%; Score 309.5; DB 2; Length 390;
Best Local Similarity 24.8%; Pred. No. 6.6e-08;
Matches 105; Conservative 76; Mismatches 192; Indels 51; Gaps 11;

Qy 825 KIOAIKMMVRLWLLGKMKNNHS--KSGTSTLRLLTTLIHSDGDLTEQGKISKPDMSRLRLAA 882
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 KVLAIVLVNRRLAAAGTTEALNIGAPIIKLKKVLWADGELSPFPKNTPKISRAVLRLTA 63

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:00:54 ; Search time 26.15 seconds
(without alignments)
4051.962 Million cell updates/sec

Title: US-09-512-581-2
Perfect score: 7193
Sequence: 1 MAHSKTRTNDGKITYPGVK.....QKGRGRPSKTPSPQPKNV 1391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2698.5	37.5	851	2 T00374	hypothetical prote
2	807	11.2	1579	2 T23142	hypothetical prote
3	632.5	8.8	1506	2 S22957	blind protein - Eme
4	571.5	7.9	1303	2 E96805	hypothetical prote
5	557	7.7	1277	2 S44451	hypothetical prote
6	309.5	4.3	390	2 T43647	hypothetical prote
7	303	4.2	780	2 F96840	hypothetical prote
8	264	3.7	990	2 H86293	protein T24D18.4 [
9	235.5	3.3	1875	2 S38173	myosin-like protei
10	234.5	3.3	852	2 T06310	hypothetical prote
11	231.5	3.2	2253	2 T30336	hypothetical prote
12	229	3.2	2663	1 S28261	nuclear/mitotic ap
13	227	3.2	1916	2 S46157	centromere protein
14	223.5	3.1	4717	2 T41581	R1F protein - yea
15	223	3.1	2139	2 T18296	hypothetical colle
16	222.5	3.1	1819	2 A71928	myosin heavy chain
17	222.5	3.1	2301	2 T02323	cag island protein
18	221.5	3.1	1927	2 G64585	nodulin-like prote
19	220.5	3.1	1727	2 T50073	cag pathogenicity
20	217.5	3.0	1020	1 QFHUH	myosin-like coiled
21	216.5	3.0	1790	2 S67593	neurofilament trip
22	210	2.9	1072	1 A37221	transport protein
23	209.5	2.9	1365	2 T45031	neurofilament trip
24	207.5	2.9	854	2 S02003	hypothetical prote
25	207.5	2.9	3488	2 T34418	neurofilament trip
26	206.5	2.9	699	2 T38073	hypothetical prote
27	206	2.9	6842	2 T29757	nucleolar phosphop
28	205	2.8	2094	2 S33124	protein UNC-89 - C
29	203.5	2.8	2364	2 A56577	tpv protein - huma
					microtubule-associ

ALIGNMENTS

RESULT 1

T00374

hypothetical protein KIAA0648 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C:Accession: T00374

R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura,

DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl

A:Reference number: Z14142; MUID:98403880

A:Accession: T00374

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-851 <ISH>

A:Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PID:g3327110

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0648

Query Match 37.5%; Score 2698.5; DB 2; Length 851;
Best Local Similarity 62.1%; Pred. No. 1.1e-121;
Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;

QY	477	LYLYATIDLNAVKALNEMWKONLLRHQVKKLLDLIKOPKTDASVKAIFSKVMVITRNL	536
DB	1	LYLYASLDPNNAVKALNEMWKONMLRSHVRELLDLHKQPTSEANC SAMFGKLTIAKNL	60
QY	537	PDGKAQDFPMKFTOVLEDEKIRKQLEVLVSPTSCQKAEQVREITKKGKGNPKQPTNP	596
DB	61	PDGKAQDFVKKFNQVLGDEKLRQLELLISFTCSKQADICVREIARLANPKQPTNP	120
QY	597	FLEMIFLERIAPVHIDFESIALIKOVNKSIDGTADDEGVPTDQAIRAGLELLKVL	656
DB	121	FLEMVKFLERIAPVHIDSEALSALVKLMKNSIEGTADDEEGVSPDTAIRSLELLKVL	180
QY	657	STHPIFSHSATFESLACLMDDKVAEALQIFKNTGSKLEEDFPHRSALLPVLHH	716
DB	181	STHTPTFSHSATFESLQCLRMEDDKVAEALQIFRNTGKHETDLPQIRSTLPIIDHQ	240
QY	717	KSKGPPRQKVAIHCHIAIFSSKTFQAFIPEFLHKSIDPSNLEHLITPLVTIGHIAL	776
DB	241	KAKRGTGHQAKVAHCHIAIFTNKEVQLAQIPEFLSRSLNADVPEQLITPLVSLGHIS	300
QY	777	APDQFAAPKWSWATFIVKDLLMNDRLPGKTKTLWVPDEEVSPTVMVKIQAIKMMVRWL	836
DB	301	APDQFASPMKSVVNFIVKDLLMNDRSTGEKNGKLSWSPDEEVSPEVLAKVQAIKLVRL	360
QY	837	LGKNNHNSKSTLTLLRTTLILHSDGDLTEQGIKISKIPDMSRLRLAAGSAIVKLAQPCYH	896
DB	361	LGKNNQSKSANSURLLSAMLVSEGLTEQKRISKSDMSRLRLAAGSAIMKLAQPCYH	420

TRDN_RABIT STANDARD; PRT; 705 AA.
Q28820; Q28637; Q28643;
30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
TRIADIN.
TRDN.
Oryctolagus cuniculus (Rabbit).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_TaxID=9986;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (SKELETAL ISOFORMS).
STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;
MEDLINE=93286104; PubMed=7685347;
Knudson C.M., Stang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;
"Primary structure and topological analysis of a skeletal muscle-
specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
J. Biol. Chem. 268:12646-12654(1993).
[2]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (SKELETAL ISOFORMS).
TISSUE=Skeletal muscle;
MEDLINE=94298946; PubMed=8026576;
Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;
"Structural diversity of triadin in skeletal muscle and evidence of
its existence in heart";
FEBS Lett. 348:17-20(1994).
[3]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (CARDIAC ISOFORMS).
TISSUE=Heart muscle;
MEDLINE=96132942; PubMed=8550602;
Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
"Biochemical characterization and molecular cloning of cardiac
triadin";
J. Biol. Chem. 271:458-465(1996).
-1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTERIN TO THE
JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
COUPLING WITH THE RYANODINE RECEPTOR.
-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. SARCOPLASMIC
RETICULUM.
-1- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; CARDIAC ISOFORM 1/CT1,
CARDIAC ISOFORM 2/CT2, CARDIAC ISOFORM 3/CT3, SKELETAL ISOFORM
1/ST1 (SHOWN HERE), SKELETAL ISOFORM 2/ST2 AND SKELETAL ISOFORM
3/ST3; ARE PRODUCED BY ALTERNATIVE SPLICING.
-1- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U31540; AAC48496.1; -
DR EMBL; L10065; AAC31488.1; -
EMBL; U31555; AAC48497.1; -
EMBL; U34201; AAC48498.1; -
Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
Alternative splicing.
INIT_MET 0
DOMAIN 1 46 CYTOPLASMIC.
TRANSNEM 47 67 POTENTIAL.
DOMAIN 68 705 LUMENAL.
CARBOHYD 74 74 N-LINKED (GLNAC. . .) (POTENTIAL).
CARBOHYD 624 624 N-LINKED (GLNAC. . .) (POTENTIAL).
VARSPPLIC 264 285 DOYAFRCYMDIEVHGLDKLPQG -> GKGSEAAAGCFKRT
DOYAFRCYMDIEVHGLDKLPQSPAPPPSPPTQASRPPTA
LPT -> ECFILFATPQIPNRQQNDLHHCFLTKKGN
GQAFCLKGC (IN CARDIAC ISOFORM 1).
MISSING (IN CARDIAC ISOFORM 2).
VARSPPLIC 264 307
VARSPPLIC 286 705
VARSPPLIC 308 705 MISSING (IN CARDIAC ISOFORM 1).
MISSING (IN CARDIAC ISOFORM 2).

FT	VARSPPLIC	415	415	A -> E (IN SKELETAL ISOFORM 2).
FT	VARSPPLIC	416	424	MISSING (IN SKELETAL ISOFORM 2).
FT	VARSPPLIC	584	584	D -> E (IN SKELETAL ISOFORM 2 AND
FT	VARSPPLIC	585	591	SKELETAL ISOFORM 3).
FT	VARSPPLIC	645	705	MISSING (IN SKELETAL ISOFORM 2 AND
FT	VARSPPLIC			SKELETAL ISOFORM 3).
FT				SKKAKEAEVSTTKKQSPKPO -> LIATVGIWGMQMEDL
FT				TPAQYPGESSGKPNSPKPO -> LIATVGIWGMQMEDL
FT				SVTLPSK (IN CARDIAC ISOFORM 3).
SQ	SEQUENCE	705 AA;	79003 MM;	13AFID84475A1361 CRC64;

Query Match 2.7%; Score 194; DB 1; Length 705;
Best Local Similarity 19.3%; Pred. No. 0.043;
Matches 138; Conservative 99; Mismatches 249; Indels 230; Gaps 28;

QY	753	KSIDPSNLEHLIT----	PLVTIIGHIALPADQFAAPWKSQWATFIVKDLLMNDRLPGKKT	808
Db	29	KVLKRTVTEDLVTTFFSPAAWLLVIALII-----	TWSAVAV-----	MFDLVDYKNF 75
QY	809	TKLWVPDEEVSPTWVKIQAIKMMVRLLGMKNHNSGISTLRLTLILHSDGL-----		864
Db	76	SASSIAKMGSDPLKLVH-DAVEETDVIYGF-----	FSLLDIISDGDDEEDDE	123
QY	865	----TEQGKISKDMSRLRLAAGSAIVKLAQPCYHEIITLEQYOLCALAINDCYQVRQ	920	
Db	124	GDEDTAKGEIEEPLKR-----		140
QY	921	VFAOKLHKGISRLRLPLEYMAICALCAKADPVKERRAHQCLVKNINVRREYLKQHAASV	980	
Db	141	---KDIH-----	KEKIEQEKPERKIPTKVHVHKEKEKEKVKKEK	177
QY	981	EKLLSLLPEVVPYTHLLAHDPDYVQDIEQLKQKVECLWFVLEILMAKNENNSHAFI	1040	
Db	178	EK-----PEKKATHKEKLEKKEPETKTVTKEKK-----	ARTKEKIEKT	218
QY	1041	RKVVENIKQTKDAQGDDAKMNEKL-----	YTCVDVAMNIMS	1078
Db	219	KKEVGKQKQKQVAKAKEVQKTPKPEKESKETAAVSKQEQKQYAFCRYMIDIFVH	278	
QY	1079	KSTYSLESKDPVLPARFTQDPKFNSTKNYLP-PEMKSFFTTPGKPK-----	TTNVL	1131
Db	279	G----DLKPGQSPAIPP---PSPTQASRPTPALPTPEEKE---GKKAEEKVTTETK	327	
QY	1132	GAVKNPLSSAGKQSQTKSSRMETVSNASSNSPSSPRIKGRLDSEMDHSENYDTWSS	1191	
Db	328	KKAEK---EDAKKSE--KETDIDMKKEPKGKSPDTKPTVK-----	VTT	367
QY	1192	PLPGKSKDRKDDSLVRSLE-ERPRGRKK-----	TPVTEQEKLGMDLTKLVQEQ	1241
Db	368	QAATKDEKEDSKAKKPAEQKGGKQKKEHEEPKSTKKEHAAPSEKQAKAKIER	427	
QY	1242	KPGSQRSRRKRGHTASESD-----EQOWPEEKR-----	LKEDILNEDEQNSPPKKGK	1289
Db	428	KEEVSAASTKKAHPAKKEEXTTKTVEQETRKQKPGKISSVLKDKKELTKKEVKVPASLKE	487	
QY	1290	GRPPKPLGGTPEEPTMTSKK-GSKKSGPPAPPEEEEEERQS-----	GNT	1337
Db	488	KGSETKK-DEKTSKPEQIKKEKPGKQKVPKPPQPPQKQKKEKPEQDQIMKPEKTHLHGP	546	
QY	1338	EOK-----	SKSKOHRVSRRAQRAESPSSASTESTQSTQPKGRGPRSKTPS	1383
Db	547	BEKVLKQVAVTTEKHVKVPKPAKKAHQEKEPPSIKTKDKPKST---SKGMEPTVES	599	

Search completed: September 25, 2001, 10:05:56
Job time: 182 sec

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X66397; CAA47021.1; -
EMBL; Y00672; CAA68681.1; -
PIR; S00928; S00928.
MIM; 189940; -
Heptad repeat pattern; Coiled coil; Proto-oncogene;
Chromosomal translocation; Nuclear protein; Transport.
DOMAIN 78 360
DOMAIN 422 571
DOMAIN 575 628
DOMAIN 758 805
DOMAIN 834 869
DOMAIN 934 979
DOMAIN 1004 1064
DOMAIN 1138 1166
DOMAIN 1196 1241
DOMAIN 1262 1304
DOMAIN 1354 1434
DOMAIN 1476 1595
DOMAIN 1833 1836
DOMAIN 1957 1964
DOMAIN 2295 2298
SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 2.8%; Score 205; DB 1; Length 2349;

Best Local Similarity 18.6%; Pred. No. 0.065;
Matches 297; Conservative 258; Mismatches 582; Indels 460; Gaps 71;

QY 19 VKETSDKTSKEEMV-----RLKMMVVK-----TFMDMQ-----DSEEEKELYLN 58
DB LQEDTDKANKOSSVLERNRNRMEIQKLSQIRVLLMLEEARGNHVIRDEEVSSADIS 527
QY 59 LALHLASDFLLKHPG-----KDVRLVACCLADIFRIYAEAPYTPDKLKDIFMFI 112
DB SSSEVISQHLVSYNIEELQOQNRLVA--LRELGETREEREQETTSKITEQLKLES 585
QY 113 OLKGLDTPKSPQFNRYFYLENIAMVKSNYICFELEDSENIETQLYRTLFVINGNHQK 172
DB 586 ALTELEQLR-----KSRQHQMLVDSIVQRDMYRILLSQ--TTGVAIP 627
QY 173 VHMHVMDLMS--SIICGDTVSQELDTVLVNLVPAHKNLKNQAYDLAKALKRRTAQAI 231
DB 628 LHASSLDVSLASTPKRSTSTVTPAPVPIESTEAI-----AKAALKQLQEIPEN 681
QY 232 YITFFNOVLMLGKTSISDLSEHPDILLILELYNIDSHLLSVLPQLEPKLS-----NDNE 287
DB 682 YKKEKAENE--KIONSOLEKLOEQVTDLSRQNTKIST-----QLDFASKRYEMLQDNV 732
QY 288 E--RLQVY-----KILA-----KMGAKDSELASONKPLWOCYLGRENDIHVPIRL 332
DB 733 EGYRREITSLHNRQKLTATQKQOIIINTMTQDLRGANEKL-----AAVEVRAE 782
QY 333 CVKFASHCLMHPDLAKDLTYLKVRSHPDPEAIRHDVIVSIVTAARKDILLVNDHLLNF 392
DB 783 NLK-----KEKEMKLSEVLRSQQRESL-----LAEORGKLLLTNLQTIQ 824
QY 393 VRERTLDKRWVRKEMGLAQIYKYALQSAGDAKAKOIAWIKDKLLHIYYQNSIDR 452
DB 825 ILER-----SETETKRLSSQIEK--LEHEISHLKKKL-----ENEVEOR 862
QY 453 LLVERIFAQYMPVHNLET-----TERMKCLYLYATLDLNAVKALEMNKKONLLRHQVKD 508
DB 863 -----HILTRNLVDQLDITKQ-----LDTETNLHLN-----TKELLNAQKE 900

QY 509 LLDLIKQPKTDASVKAIFSKVMVITRNLDPGKAQDFMKK-----FTQVLEDEKIRKQL 563
DB 901 IATL--KQHLN-----MEVQASQSSQRTGKGQPSNKEDVDLVSQLRQTEEQVNDLK 952
QY 564 EVLVSPSTCSCKAECVREITKKGKLNPKQPTNPFLEMIKFLERAPVHIDTESIALIK 623
DB 953 ERLKTSTNVQOYQAMVTSLEBSLNKEKQVTEEVKRNIEVRLKESAEFOTQLEK-----KLM 1009
QY 624 OVNKSIDGTADDEGVPTDQAIRAGLELLKVLSTHPISTFHSATFESLLACLKMDDEK 683
DB 1010 EVEKEQELQDDKRAI-----ESMEQOLSELK----- 1037
QY 684 VAEALQIFKNTGSKIEDPHIRSAIPLVLHKKKGGPPROAKYAIHC----- 732
DB 1038 -----KTLSSVQNEVEQALQRASTALS--BQARRDCQEQAKTAVEAQNKYERELMLH 1089
QY 733 ---THAIFSSKE--TQFAQIFELHKSIDPSNLEHLITPLVTIGHIALLAPQFAPW-- 785
DB 1090 AADVEALQAQAEQVSKMASVRQHLEETQKAESQLL-----ECKASWEE 1133
QY 786 -----KSWATFIVK--DLLMNDRLPGKRTTKLWVPDDEVSPEVMVKIOAKMMVRWLIG 838
DB 1134 RERMLKDEVSKVCRCEDLEKQNRLLHQIEKL---SDKV-----VASVKEGVQGPLN 1183
QY 839 MKNHKSSTGLRLTLTILHSDGLTEQGIKSKPDMSKRL-----LAAGSAIVKLAQ 891
DB 1184 V-----SLSEEGSQEQILEILRIRREKEIAETRFEVAQVE 1220
QY 892 EPCYHEITLBOYLQCAL--AINDECYQVRQVFA-----QKLHKGSLRLPLPLEYMAIC 943
DB 1221 SLRYRQVRELLERELOELEDLSNAEREKV-QVTAKTMAQHEELMKTKTETMNVME----- 1274
QY 944 ALCADKDPVKERRAHARQCLVK--NINVRREYL-----KOHAAVSEKL-----LSLLPEY 990
DB 1275 ---TNKMLREBEKERLEQDLQMQAKVKLELDILPLQEAANESEKSGMLQAEKILLEED 1331
QY 991 VVPYTI---HLLAH--DDPYVKVQDIEQLKDV--KECLWFVLEILMAKNE-----NN 1035
DB 1332 VKRWKARNQHLVSOQKDPDTEYRKLSKEVHTKRIQOLTEEIGRLKAEIARSNASLTN 1391
QY 1036 SHAFIRKRVENIKOTK---DAQGPDDAK---MNEKLYTVCDVAMNIMSKSTTYSLESP 1088
DB 1392 NONLIQSLKEDKNVRYTEKETIQKDLDAKIDIQEKVITITQVKKIGRYKTYEELKQ 1451
QY 1089 KDPVLPARFETTPDKNFNTKNYLPPEMKSEFTPGKPTTNVLGAV---NKPPLSAGKOS 1145
DB 1452 QDKVNETSAQSSGDHQEQHVSVQEMQELKETLQNAETKSLSQVENLOKTLS--KET 1509
QY 1146 OTKSSRMETVSNASSSSNPSPSGRIKGRLDSDSEMDSHSENYTMSPLPGKSKDXDRDSD 1205
DB 1510 EARNLQEQTVQLQSLS-----RLRQDLQDRTTQEE--QLRQOITEKEEKTRKATV 1558
QY 1206 LVRSLEKPRGRKKTPTVTEQEEKL-----GMD-----LTKLVQEKPKGSQSRK- 1251
DB 1559 AAKSIAHLAG--VKDQLTKENEELKQKRGALDQKDELDVRITALKQYEGRISRLERL 1617
QY 1252 RGHASEDEQOWPEEKRLKEDILENEDEQNSPPKKGKRG-----RPP-----KPLCGGTP 1302
DB 1618 REHQERHLEQDEPOEPSNKVPEQOQITLKTTPASGERGIASSTDPPTANIKP---TP 1673
QY 1303 -KEEPTMTKSKGSKKKSGP-----PAPEREEERQ----- 1333
DB 1674 VVSTPSKVTAAMAGNKSTPRASIRPMVTPATVNTPTTPTATVMTPTTQVSEQAMQSEG 1733
QY 1334 -----SGNTEQKSKQHEVSR---RAQORASP----- 1359
DB 1734 PVEHVVPFGSTSGSVRSRSTPNVQPSIQPILTVQOQTATAPVQPTQOSHPOIEPANQEL 1793
QY 1360 ESSAIESTQSTPQKGR-----GRPSKTPSPSQPK 1389
DB 1794 SSNIVEVQSSPVERPSTSTAVFGTVSATPSSSLPKR 1830

FT	DOMAIN	204	382	INTERACTS WITH RPA194.				
FT	DOMAIN	384	587	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).				
FT	DOMAIN	601	617	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).				
FT	MOD_RES	563	563	PHOSPHORYLATION (BY CK2) (BY SIMILARITY).				
FT	VARSPLIC	241	241	K -> KWVITTSVRAE (IN ISOFORM BETA).				
FT	CONFLICT	3	3	D -> A (IN REF. 2).				
FT	CONFLICT	133	133	R -> S (IN REF. 2).				
FT	CONFLICT	291	292	YA -> SV (IN REF. 2).				
FT	CONFLICT	456	456	S -> P (IN REF. 2).				
FT	SEQUENCE	699 AA;	73720 MW;	DFD4AD94EDF659FB CRC64;				
Query Match		2.9%;	Score 206.5;	DB 1; Length 699;				
Best Local Similarity		22.3%;	Pred. No. 0.011;					
Matches 93;	Conservative 55;	Mismatches 172;	Indels 97;	Gaps 13;				
QY	1045	ENTKQTKDAQGPD	AKMNEKLYTVC	DAVAMNI	I	MSKSTYSLESPKDPVLPARFFTPQDKN	1104	
DD	228	DSEEEKAATPKK	TVPKKQVAKAP	KAATTPTRKSSSESSSEDEEEQK	---	KPMKN	284	
QY	1105	FSNKTNYLPMKSF	TPGPKPTTNVL	GAVKNKLSSAGKQ	SQTKSSRMETVSNASSSN	-	1163	
DD	285	KGCPYSAPP	-----	PSAPPKKSIG	-TQPKKAVEKQPVES	-	EDSSDESDSSSE	334
QY	1164	-----	-----	PSSGRIGRLDS	SEMHSNEDYTMSSPLPKPKSKDKRDS	D	1205	
DD	335	EKKPPTKAVSVK	ATTKPPAKKA	ESSSDSDSSE	-DDEAPSKPAGTTKNSNKP	PAV	393	
QY	1206	LYRSELEKPRGRK	TPV-----	-----	TEQEKLGMDLT	-KLVOEQKPKGSQ	1249	
DD	394	TTKSPAVKAAAP	KQVPVGGQK	LLTKRADSSSESSSEEESEKTKMVA	TKPKATAK	453		
QY	1250	RKRGHTASEDEQ	OWPEEK	---L	KEDILNEDEONSPKKGKGRPPKPLGGG	TPV	1302	
DD	454	-----	-----	ALSLPAKQAQGS	RDSSDSDSSSESEEEKTSKAVKKKQK	VAGGAPSKPA	507	
QY	1303	-----	-----	KEEPMTKSKKSGK	KKSGPPA-----	-----	PEEE	1328
DD	508	SAKKGKAESSN	SSSDSDSDSS	DEEKLKGKSPRQAPKANGTSALT	AQNGKAANKSEEE	567		
QY	1329	EEEROSGNTQK	SKSOHRYSV	RAQRAQRAQSP	ESSAIE	-STOSTPQKGRGRPSKTPSP	1384	
DD	568	EKKKAADVVS	KSGLKKRQNE	AKAEATPQAKIKLQTPNTP	PKRKKKEGRASSP	624		
RESULT 9								
GCG4_HUMAN								
ID	GCG4_HUMAN	STANDARD;	PRT;	2230	AA.			
QI	Q13439;	Q14436;	Q13270;	Q13654;				
DT	01-OCT-2000	(Rel. 40,	Created)					
DT	01-OCT-2000	(Rel. 40,	Last sequence update)					
DT	01-OCT-2000	(Rel. 40,	Last annotation update)					
DE	GOLGI AUTOTANTIGEN,	GOLGIN SUBFAMILY A 4	(TRANS-GOLGI P230)	(256	KDA			
DE	GOLGIN)	(GOLGIN-245)	(72.1	PROTEIN).				
GN	GOLGA4.							
OS	Homo sapiens (Human).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
OX	NCBI_TaxID=9606;							
RE	[1]							
RE	SEQUENCE FROM N.A.							
RE	MEDLINE=96215236;	PubMed=8626529;						
RE	Erlich R.; Gleeson P.A.; Campbell P.; Dietzsch E.; Toh B.-H.;							
RE	"Molecular characterization of trans-Golgi P230: a human peripheral							
RE	membrane protein encoded by a gene on chromosome 6p12-22 contains							
RE	extensive coiled-coil alpha-helical domains and a granin motif.;"							
RE	J. Biol. Chem. 271:8328-8337(1996).							
RE	[2]							
RE	SEQUENCE FROM N.A.							
RE	Seelig H.P.;							
RE	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.							
RE	[3]							
RE	SEQUENCE OF 131-2230 FROM N.A.							

Qy	681	---	DEKVAEALQIFKNNTGSKIBEDPPHIRISALLPVLHHKS-----	-	K719								
Dd	873	QELQEKTRVQERL--	NEMBLKEQLERNRDSPLQTVEREKTLLITEKLQOTLEEVKTLTQE	930									
Qy	720	KGPRQAAYAIHCIIHALFSKETQFAIEPLHKSLDPSN----	LEHLITPLTVIGHITA	774									
Dd	931	KDDLKOLOESLQI-----	ERODKLSDIHDVTNMNDTQOEURLNALESUKHOETINTL-	983									
Qy	775	LLAPDQFAAPKWISWATFIVKDLMLMDRLPGKTKTKLWPDPBSEVPETVKVIQAIKMVMVR	834										
Dd	984	-----	KSIISEVSRLHM-----	EENGETKDPEQ-----	Q101								
Qy	835	WLGLMKNHNSKSTSTLRLLTTILSHSDGLTEQGKISK--	PDMSRRLRLAAGSAIVKLAAQ-	891									
Dd	1011	KWVGIDKKQBLEAKNT-OTLTADVADKNETIEQQORKIFSLOEQNELQOQMLESVIAEKEOL	1066										
Qy	892	----	EPCYHEIITLBOYLICALAINDECYOYOVPAOKLHGCLSLRLPLEYMAICALC	946									
Dd	1070	KIDLKENIEMTIENQBELRL----	LGEDELKQOEIVAQEKNHAIRK--	EGELSRTCDBL	1122								
Qy	947	A--	KDPKERRAHAROCCLKVINVRREYLKHAQAAVSE-----	KLILLSLPEYVVPVTI	996								
Dd	1123	AEVEELKESQOLQEQKQOLLNVQEMSEMOKKINEIENLNKELNKELTLEHMETERL	118										
Qy	997	HLLAH-DPDYKVYVODI-EOLKDVYKCELFVLEITLMKAENNENSHAFIRKVMVENIKOTKDAQ	105										
Dd	1183	ELAQLKNENYEYKSIKTERKVLKE-----	LQKSPETERDHLRGYIREATEATGLQTKBEL	123									
Qy	1055	GPDDAKNEKLYIVCDVAMNIINMSKSTYSLESKPDPVLPAREFTQDPKNFNSTKNYLP	111										
Dd	1238	KIAIHILKHEQETIDELRRSV--SEKT-----	AQIINTQDLEKSHTK--	LQE	128								
Qy	1115	EMKSFFTPGPKPTNYLVGAVNKRPLSSAGKOSOTKS--	SRMETVSNASSSNSPSSPCRIGK	117									
Dd	1281	EIP-----	VLHQEOELLPNVKVSETQETMNELELLTQSTTKOSTTTLARI--	132									
Qy	1173	RLDSSMDHSE-NEDYTMSPLPGKSKDRDDSDLVRSLELERPRGRKKTPTVTE-----	122										
Dd	1327	-----	EMERLRLNEKEFOEQEETIKSLTKERDNLKTIKEALEYKHVDQKHEHRETAKIQE	138									
Qy	1225	----	QBEKLGQ-----	DDLTKLVQEQ-----	KPK-----	GQSQRKRCH-----	TA	125					
Dd	1382	SQSQEQSGLSMKEKDNETTKIVSEMDFQPKDSALLRIEIMLGLSKRLQESHDEMKSVA	144										
Qy	1257	SESDEQ-----	WPEKRLKEDI-----	LENEDQONSPPKPKGRGRP-KPLGGGTP	130								
Dd	1442	KEKDDIQRLOEVLOSSEDQOLKENIKIIVAKHLEETEELKVAHCCLKEQETTNELRVNLS	150										
Qy	1303	KEPTMTKTSK-----	GSKKKSGPPAPEEEEE--	EROSGNTQEKSKQHVRRAQQ	135								
Dd	1502	EKETEISTIQOLEAINDKLONKIQIEKEEQNLNKQISEVOENVNELKFQEHKAKD	156										
Qy	1355	RA-ESPRESSAIST	1367										
Dd	1562	SALQSIESKMLELT	1575										
RESULT 4													
ID	RIFI_YEAST	STANDARD;	PRT; 1916 AA.										
AC	P29539;												
DT	01-APR-1993 (Rel. 25, Created)												
DT	01-OCT-1994 (Rel. 30, Last sequence update)												
DT	30-MAY-2000 (Rel. 39, Last annotation update)												
DE	RIFI PROTEIN (RAP1-INTERACTING FACTOR 1).												
GN	RIFI OR YBR275C OR YBR1743.												
OC	Saccharomyces cerevisiae (Baker's yeast).												
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;												
OX	NCBITaxid=4932;												
RP	[1]												
RP	SEQUENCE FROM N.A.												

[illegible]


```
174 --ILGEVISEDSVPLEVRLIFNFKFLYNPEIPEGLNVTSDCGYEVSLILCDITYSNRM 231
230 EPIYITFFNQVWML-----GKTSIDSLSEHFVDFDLELYNIDSHLLSLLOLEFKKSN 284
232 SRHLTKYSEIHEATNDNNSRLTVVVKHLKVLRLWEIVPELINAVIGFIHELSE 291
285 DNEERLQVVKLAKMGA-KOSELASQNKPLQCVYLGRENIDHVPRLCEVKFASHCLMN 343
292 NELFKRKAATLGITLTSTYSOLNFVTSHTDFFKAWISKIADISPDVRVWETESIPQIIAT 351
344 HPDLAKDLTEYLKVRSHDPEAIRHDVIVSI-----VTAACKDIL--LVNDHLLNFVRETT 397
352 REDISKELNALAKTFFIDSDPRVRTSVIMFNKVPVTEIWNKNTKALYTSLLHARE-- 409
398 LDKRWRVRKEMGMAQIYKRYALQ--SAAGKDAKAQIAWKDKLLHYIYONSIDDRLL 454
410 --KHKEVELCINTMAKYSNSLNEIERTYONKEIWEIIDIPSTLYNLYINDLNINEQ 467
455 VERIPAQVWPHNLETTERMKCLYLYATLDLNAVKALNEMKQNLRLHQVKDLDLILK 514
468 VDSVIFEYLLPFPDNRKVRHLLTVLSHFDFKAPTSPFAFNARQIKISFAISKYIDFSK 527
515 QPKTDAVKAFSKVMV-----ITRNLPDCKAQDFMKKFTOVLEDDKIRKQLEV 565
528 FLNQESMSSGQPLVMNKYNOTLQWLASGLSDSKAIDALETIKQF--NDRIFPYLNA 585
566 LVSPTCCKQAGCVREITTKLGNPK-----QPTNPFLEMIKFLLEIRIAPVH 612
586 CVTNDIPLETFCNVCYNELVSKLOTPGLFKYINISGASIMPRD-IAKVIOILLFRASPII 644
613 IDTESIALIKOVNKSIDGTADDEDEGVPTQOATRAGLE--LLKVLSTHPISEF-HSAET 669
645 YVNSVINSVLLNSNNS-----DAKQLDKRRLDIDSKVNPFTLFDQDQRT 689
670 FESLLACL-----KMDKEVAEAAQIFKNTGS-KIEEDPPIRSALLPVLHKKSKGP 722
690 LKTIKLDLDDPAEKNDNLSEALKTIYKASKTLKDQVDD--TFFTKLIDPFAVESK 747
723 PROAKYAIHCIAIFSSKETQFAQI---FEPLHKSLDPSNLEHLTPIVLTGHTALLAPD 779
748 PEITKYATKLI-ALSPKAEETLKIKIRILPDLQKQKYFTSHIIVLMFEIKKPPHVLND 806
780 QFAAPKSWATFVVKLLMDRLPLGKTKTL-WVPDEVSPETWV-----KIQAIKMVR 834
807 D-----STDIISYLIKEVLLSNQVVGDSKKEIDWVEDSLSDTKYSATGNVFTLKLFTN 861
835 WLGM-----KNHNSKST-STLRLLTTLHSDGDL-TECGKISKPDMS-----RLRLAAG 883
862 KIRSTAPVPRDELAESTETKMLFFLYIASGGELISEFKNKEFYPTPSNTQTKLRVAG 921
884 SAIVKLAQPCYHEIITLEQYOLCALAINDECYQVQVFAQKLHKLGRRLRPLEYMAIC 943
922 IQVLKLARISLNNFKPSDIKILNVLDEDESLPVRFTLEQKDYVANELISIKFLPLV 981
944 ALCAKADPKVERAHARQCLVINVRREYLKQHAAYSEKLLSLPVEYVPTIHLHADP 1003
982 FFTAYEPDVELTKTKIWI--NFTFGLSKFKK-----GTIFERALPLRIHAIAHP 1030
1004 DYVKVQDIE-----QLKDXKCELMFVLEILMAKNENNSHAFIRKVMENIK-----OTK 1051
1031 DIVGGLDSEGDAYLNALTAIDYLLFYFDSIAAQENFSLLYLSERVKNYQDKLVEID 1090
1052 DAQGDDAK-----MNEKIYTVCDVAMNIMS-----KSTYS-----LESPPKDPVLP 1095
1091 EEEGPQKEEAPKHPYQOKMYIIGELSQMILLNLKELKKNQHSAYPGKLNLPDLFPK- 1149
1096 REFTQPKNFNTKNYLPPEMKSFPTPKPTTYNVLGAVNKLPSAGQSOTKSRMETV 1155
1150 -FATVQEAQLS-FKTYIPESIT-----EKTONNIKAKIGRILHTSQTORLOKRLAAH 1201
1156 SNASSSSPSPGRIKGLDSEMDHSENEDYTWSSPLPGKSD-KRDDSDILVRSELEKP 1214
```

```
Db 1202 ENNESQKKKKKVVHHARSQAADDEEGDRESDDSDSYSPSNKNETKKGHENIV---MKKL 1258
Qy 1215 RGRKKTPVTEQEEKLGMDL 1234
Db 1259 RVRKEVDYKDDDED---DDI 1274

RESULT 2
MLP1_YEAST
ID MLP1_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C.
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L01992; AAA34783.1; -;
CC EMBL; X73541; CAA51948.1; -;
CC EMBL; Z28320; CAA82174.1; -;
CC PIR; S38173; S38173.
CC SGD; S0001803; MLP1.
CC MYOsin: Heptad repeat pattern: Coiled coil; DNA repair.
CC DOMAIN 69 487 COILED COIL (POTENTIAL).
CC FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
CC FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
CC FT CONFLICT 301 301 R -> A (IN REF. 1).
CC SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 3.3%; Score 235.5; DB 1; Length 1875;
Best Local Similarity 19.5%; Pred. No. 0.002;
Matches 301; Conservative 246; Mismatches 602; Indels 391; Gaps 72;

Qy 19 VKESDKTSKE-----EMVRLKMWKVTMDMDODSEEEKEL-VLNALHLASDFFLKH 71
Db 278 IKGUSDLSNKSQBSAEFMSLQKRLVDLLESOLNAVKRELNSIRELNTAKVIADSKKQT 337
Qy 72 PG-----KDVRL---VACCLADIFRI-----YAPEAPYTPDKLKDIFMTTROLKGL 118
Db 338 PENEDLLKELQITKEKLAQCEKLECLRSSITDEADEDENLSAKSSSDIFILKQLIKER 397
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:02:54 ; Search time 16.41 Seconds
(without alignments)
2903.680 Million cell updates/sec

Title: US-09-512-581-2
Perfect score: 7193
Sequence: 1 MAHSKTRTNDGKTYPPGVK.....QKGRGSPKTPSPSQPKKNV 1391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	557	7.7	1277	1 YMW6_YEAST	Q04264 saccharomyc
2	235.5	3.3	1875	1 MLP1_YEAST	Q02455 saccharomyc
3	229	3.2	2663	1 CENE_HUMAN	Q02224 homo sapien
4	227	3.2	1916	1 RIF1_YEAST	P29539 saccharomyc
5	217.5	3.0	1020	1 NFH_HUMAN	P12036 homo sapien
6	217	3.0	1790	1 USOL_YEAST	P23886 saccharomyc
7	212	2.9	831	1 NFH_RAT	P16884 rattus norv
8	206.5	2.9	699	1 NP14_HUMAN	Q14978 homo sapien
9	205.5	2.9	2230	1 GOG4_HUMAN	Q13439 homo sapien
10	205	2.8	2349	1 TPR_HUMAN	P12270 homo sapien
11	202	2.8	2748	1 NUM1_YEAST	Q00402 saccharomyc
12	199	2.8	2464	1 MAPB_MOUSE	P14873 mus musculu
13	195.5	2.7	1679	1 Y109_YEAST	P40457 saccharomyc
14	195	2.7	488	1 CYL2_BOVIN	Q28092 bos taurus
15	194	2.7	705	1 TRDN_RABIT	Q28820 oryctolagus
16	194	2.7	1533	1 TP2A_PIG	Q46374 sus scrofa
17	194	2.7	1539	1 Y373_HUMAN	Q15078 homo sapien
18	193.5	2.7	3210	1 CENF_HUMAN	P49454 homo sapien
19	192.5	2.7	2469	1 TEGU_HSVSA	Q01056 herpesvirus
20	190.5	2.6	2468	1 MAPB_HUMAN	P46821 homo sapien
21	188.5	2.6	633	1 MLH_TETH	P40631 tetrahymena
22	188.5	2.6	1453	1 Y373_BOVIN	Q9tu23 bos taurus
23	188	2.6	1841	1 CC12_SCHPO	Q10059 schizosacch
24	187.5	2.6	2017	1 MYSN_BROME	Q93323 drosophila
25	187	2.6	1087	1 NFH_MOUSE	P19246 mus musculu
26	186.5	2.6	1130	1 YL17_CAEEL	Q11102 caenorhabdi
27	186.5	2.6	2649	1 BPAL_HUMAN	Q03001 homo sapien
28	186	2.6	728	1 TRDN_HUMAN	Q13061 homo sapien
29	186	2.6	2869	1 RBP1_PLAVB	Q00798 plasmodium
30	185.5	2.6	2364	1 SPOC_HUMAN	Q10082 homo sapien
31	184.5	2.6	3678	1 DMD_MOUSE	P11531 mus musculu
32	184.5	2.6	3924	1 ANK2_HUMAN	Q01484 homo sapien
33	184	2.6	1466	1 SPA2_YEAST	P23201 saccharomyc

ALIGNMENTS

RESULT 1

YMW6_YEAST

ID YMW6_YEAST STANDARD; PRT; 1277 AA.

AC Q04264; Q04780;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE HYPOTHETICAL 147.0 KDA PROTEIN IN ABP2-CHL12 INTERGENIC REGION.

GN YMR076C OR YM9582.01C OR YM9916.15C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE OF 1-682 FROM N.A.

RC STRAIN=S288C / AB972;

RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 659-1227 FROM N.A.

RC STRAIN=S288C / AB972;

RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

	XX	24-APR-2001 (first entry)	
	DT	Oligonucleotide D1954.	
	XX	Electron-transfer group; ETM; mismatch; genotyping;	
	XX	gene expression; ss.	
	KW	Synthetic.	
	OS	WO200107665-A2.	
	PN	01-FEB-2001.	
	PI	26-JUL-2000; 2000WO-US20476.	
	PF	26-JUL-1999; 99US-0145695.	
	XX	17-MAR-2000; 2000US-0190259.	
	PR	(CLIN-) CLINICAL MICRO SENSORS INC.	
	PA	Umek RM;	
	XX	WPI; 2001-159728/16.	
	XX	Nucleic acids containing electron-transfer group, useful as labels in	
	DR	hybridization assays, e.g. for genotyping, allowing repeat analyses on	
	XX	a single surface -	
	PT	Example 6; Page 127; 159pp; English.	
	PS	The present invention relates to a composition comprising two nucleic	
	XX	acids each containing an electron-transfer group (ETM) having	
	CC	different redox potentials. The invention is used for electronic	
	CC	detection of nucleic acids, especially of substitutions (mismatches)	
	CC	and single-nucleotide polymorphisms, e.g. for genotyping,	
	CC	monitoring gene expression.	
	XX	Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;	
	QQ		
	SS		
	TT		
	UU		
	VV		
	WW		
	XX		
	YY		
	ZZ		
	AA		
	BB		
	CC		
	DD		
	EE		
	FF		
	GG		
	HH		
	II		
	JJ		
	KK		
	LL		
	MM		
	NN		
	OO		
	PP		
	QQ		
	RR		
	SS		
	TT		
	UU		
	VV		
	WW		
	XX		
	YY		
	ZZ		
	AA		
	BB		
	CC		
	DD		
	EE		
	FF		
	GG		
	HH		
	II		
	JJ		
	KK		
	LL		
	MM		
	NN		
	OO		
	PP		
	QQ		
	RR		
	SS		
	TT		
	UU		
	VV		
	WW		
	XX		
	YY		
	ZZ		
	AA		
	BB		
	CC		
	DD		
	EE		
	FF		
	GG		
	HH		
	II		
	JJ		
	KK		
	LL		
	MM		
	NN		
	OO		
	PP		
	QQ		
	RR		
	SS		
	TT		
	UU		
	VV		
	WW		
	XX		
	YY		
	ZZ		
	AA		
	BB		
	CC		
	DD		
	EE		
	FF		
	GG		
	HH		
	II		
	JJ		
	KK		
	LL		
	MM		
	NN		
	OO		
	PP		
	QQ		
	RR		
	SS		
	TT		
	UU		
	VV		
	WW		
	XX		
	YY		
	ZZ		
	AA		
	BB		
	CC		
	DD		
	EE		
	FF		
	GG		
	HH		
	II		
	JJ		
	KK		
	LL		
	MM		
	NN		
	OO		
	PP		
	QQ		
	RR		
	SS		
	TT		
	UU		
	VV		
	WW		
	XX		
	YY		
	ZZ		
	AA		

Qy	491	ttaataacaatg	gccacaatcagaagagtc	ccatatagcacatggtagaccttatgagctctata	550
Db	366	#####	#####	#####	307
Qy	551	ttatttgaagg	tgtacacagtgctc	caggagcgttttttggtacacggttttttagtaaatcctgg	610
Db	306	#####	#####	#####	247
Qy	611	tacctgctcata	agaatttaaacacagc	acatgatgatttggc	670
Db	246	#####	#####	#####	187
Qy	671	ggacagctcaag	ctattgacocatata	taccaccttttttaacacagggttctgctgattg	730
Db	186	#####	#####	#####	127
Qy	731	ggaaacaatcat	cagcgatttgc	acagacatgctcttggacttaatttggagcctctaca	790
Db	126	#####	#####	#####	67
Qy	791	atttgatagtc	attgctgctctctgtt	taccacagcgttgatttaaat	850
Db	66	#####	#####	#####	7
Qy	851	atgata	856		
Db	6	#####	1		
RESULT 13					
AAF58259/C					
ID	AAF58259 standard; DNA; 936 BP.				
AC	AAF58259;				
DT	24-APR-2001 (first entry)				
DE	Oligonucleotide D2004.				
KW	Electron-transfer group; ETM; mismatch; genotyping;				
KW	gene expression; ss.				
OS	Synthetic.				
PN	WO200107665-A2.				
PD	01-FEB-2001.				
PF	26-JUL-2000; 2000WO-US20476.				
PR	26-JUL-1999; 99US-0145695.				
PR	17-MAR-2000; 2000US-0190259.				
PA	(CLIN-) CLINICAL MICRO SENSORS INC.				
PI	Umek RM;				
WPI	2001-159728/16.				
PT	Nucleic acids containing electron-transfer group, useful as labels in				
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on				
PT	a single surface				
PS	Example 6; Page 128; 159pp; English.				
CC	The present invention relates to a composition comprising two nucleic				
CC	acids each containing an electron-transfer group (ETM) having				
CC	different redox potentials. The invention is used for electronic				
CC	detection of nucleic acids, especially of substitutions (mismatches)				
CC	and single-nucleotide polymorphisms, e.g. for genotyping,				
CC	monitoring gene expression.				
SQ	Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other:				

XX Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer;
 XX
 OS Homo sapiens.

Db 231 aaatacggaaacacaqaagtccaaaagcaaacacagcaccgaagtgtcaaggagagacacagcagaq 290

602 tcaatagaggggacagagatgatgaagaggagggtgtaagtccagatcacagctatccgt 661
1942 gcaggctctgaactgcttaagggtactctcaattacacatcccctcattctcattctgct 2001
662 tcaggacttgaaactcttaagggtctgctcttttacacatcccctcattctcattctgca 721
2002 gaacatttgaaactcttaagggtctgctctttgaataatgatgaaagtagcagaagct 2061
722 gagacatgatgctctgttacagtgccctaagaatgagagatgacaaggtagcagaagct 781
2062 gaactcaaaattttcaaaaacacaggaagcaaaatgaagagagattttccacacatcaga 2121
782 gctattcaaaattttgaataacaggtccaaatgaacagacacgtctcccagatcaga 841
2122 tcagccttgctctctgtttacacacaaatctaaaaagagccccccgctcaagccaa 2181
842 tcgaccttaattcccaattttacatacaaaagcaaaagagggggtactccacaccaagcaaaa 901
2182 tatgcattcatttgatcccatcgatattttctagtaaaagagaccagtttgacacagata 2241
902 caggctgtgcactgtatacacgcatattccacaaataaagaagtcacagcttgacagatt 961
2242 ttgagccctgcataagagccttagatcccaagcaacctggaaacatctcataaacaccattg 2301
962 ttgagccactcagtaggaggtctggaatgctgtagtgcagaaacaaacttataaccctcta 1021
2302 gttactattggtcatattgctccttgacactgacaaatttctgctccttggaatct 2361
1022 gttcattgggacacatttctgttagcaccagatcagtttcttcccacaaatgaaatct 1081
2362 tgggtgactcttcatgttgaagatcttctcatgaatgatcggttccagggaagaaag 2421
1082 gtatgacaaattttattgtgaagatctgctaataatgacaggtcaacaggtgaaag 1141
2422 acaactaaacttggttccagatgaagaagtatctctgagacaaatggtccaaatccag 2481
1142 aatggaaaactgtggtctccagatgaaggttctccctgaaatgactagcaaaaggtacag 1201
2482 gctattaaaatgatggtctgacttggaatgaaataatcacagtaaatcagga 2541
1202 gcaattaaactctgttaaggtgctgtgtgtgatgaaataacacagcttaaatctgcc 1261
2542 actctaccttaagattgttaacaaacaaatattgcatagtgatggagacttgacagaaag 2601
1262 aattcaacctctggttattatcagagatgtgtgttagtgagggtagcttgacagagcaa 1321
2602 ggaaaaattagtaaacacagatgtcagctgcagactgtcgtcgtggagtgctattgtg 2661
1322 aagagatcagtaaatctgatagtctcgtcgtcgtgattgctggtgtagtgcataatg 1381
2662 aagctggcacaagaacctgttaccatgaaatcatcacattagaaacaaatcagctatgt 2721
1382 aagctgtcagaacctgttaccatgaaatattaccaccaagaacaggttcaagctgtg 1441
2722 gcattagctatcaacgatgaatgctatcaagtaagacaagtgtttgccacagaacctcac 2781
1442 gcacttgttaataatgatgaggttaccagtaaggaagcagataattgtctcagaagctgcat 1501
2782 aaaggcccttccggtttacagcttccacttgatgatatgcaatctgtgcccctttgtgca 2841
1502 aagggcactgtgaagttactgcctccacttgagttatgagcagatcttgcctgtgtgcc 1561
2842 aaagatcctgtaaggagagaagagctcatgctagggcaatgttggtaaaaataataat 2901
1562 aaagatcctgtgaaggagagaagagcagcagcagcaaatgtttactgaaaaatacagtt 1621
2902 gtaaggcgggagtatctgaagcagcagctgagctgtgttagtgaaaaattattgtctctta 2961
1622 ataccgaggaatacacattagcagaatccctatggtcactgagaaattattatcactgtg 1681
2962 ccagagtatgtttccatatacaattcacttttggcacatgaccagattatgtcaaa 3021
1682 cctgaatatgtttccatatacattgattccactgtagcccatgacccatgacagatttacaaga 1741

QY 3022 gtacaggatattgaacaaacttaagaatgtttaagaatgtcttgggtttgttctggaata 3081
Db 1742 tcacaagatgttgatcagcttgatatacaagaagtgctatggttcattgctgaagtt 1801
QY 3082 ttaattgctaaaaatgaaataacagtcacgctctttatcatcagaagaatggtgagaaatatt 3141
Db 1802 ttaattgacaaaagatgaaacaaatgaccttctttatgagaagaatgagcagagaacatc 1861
QY 3142 aaacaaacaaagatcccaagaccagatgacaaaatgacaaaatgacaaaactgtacact 3201
Db 1862 aagttacacagagatccccagctccagatgaatccaaagacaaatgaaaactgtataca 1921
QY 3202 gtgtgtgatttgccatgaataatcatctgtaaaagagtactacatcacagtttggaaatct 3261
Db 1922 gtatgtgtgtgctctctgtgtataaataagtaaaagtcttgtgcaatgcagattca 1981
QY 3262 cctaaagaccgggtactaccagctcgtttctcactcaactcagacagaatcttcagtaac 3321
Db 1982 ccaaaaggaccagctccccaatgaaaatttttacacaaactgaaaaggactctgttaac 2041
QY 3322 accaaaattatctgctcctgaaatgaaatcatattttcactcctgaaaacacttaaaaca 3381
Db 2042 gataagatttatattcagaagagacaaagtagtctctgttaacaggaagacaaagcct 2101
QY 3382 accaatgttctagagctgttaacaagccactttctacagcaggaagcaactctcagacc 3441
Db 2102 gctggagtactagtgctgataaataagcctttatcagcaacggaaggaacccctatgtt 2161
QY 3442 aatatcatcagaatggaacctgtaagcaatgcaagcagcagctcaaatccaaagctctcct 3501
Db 2162 agaagcactggcactgagactggaagcaatattaatgaaattcagagctgaacccctta 2221
QY 3502 ggaagaataaaggagggtgattgattgtgtaattggaatgagatcacagtgaaaatgaaga 3557
Db 2222 accygaatctgataaggggaagcagaggttcagagcgacagaacactggagttagtga 2277

RESULT 4

AAV87629

ID AAV87629 standard; cDNA; 439 BP.

XX AAV87629;

XX 12-FEB-1999 (first entry)

XX EST clone DY17.

DE Homo sapiens.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;
chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

XX WO9845437-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-0506956.

XX 10-APR-1997; 97US-0837312.

XX (GEM) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-070078/06.

XX New polynucleotides encoding human secreted proteins - derived from
e.g. human blood, kidney, foetal lung, placenta, testes, brain,
ovary, pituitary, retina and colon cDNA libraries

CC cell transfected with the expression vector, and a method for producing
 CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
 CC cytosolic activity, and acts to suppress cell proliferation. The AS3
 CC gene is useful as a marker for the efficient diagnosis and treatment of
 CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
 CC antisense agents for sequence specific modulation of gene expression. The
 CC AS3 protein may be used in the treatment of disorders caused by aberrant
 CC modification or mutation of a gene encoding an AS3 protein, misregulation
 CC of the AS3 gene or aberrant post-translational modification of the AS3
 CC protein. This sequence represents the human AS3 cDNA sequence with an
 CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
 CC compared with the claimed AS3 cDNA sequence AAA28051.
 XX
 SQ Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 4173; DB 21; Length 5355;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctcattcaaaactagaccatgaggaataattacatatccgctgggtcaag 60
 Db |||||
 QY 61 gaatatcagataaaatatctaaagaggagatggtgagacgattaaagatggtgaaa 120
 Db |||||
 QY 210 gaatatcagataaaatatctaaagaggagatggtgagacgattaaagatggtgaaa 269
 Db |||||
 QY 121 actttttagtatgagaccagactctgaagagaagaaaggagctttatttaaacctagct 180
 Db |||||
 QY 270 actttttagtatgagaccagactctgaagagaagaaaggagctttatttaaacctagct 329
 Db |||||
 QY 181 ttacatcttgcctcagatctttttctcaagcatcctggttaaagatggtcgttactggta 240
 Db |||||
 QY 330 ttacatcttgcctcagatctttttctcaagcatcctggttaaagatggtcgttactggta 389
 Db |||||
 QY 241 gctctgctctgctgatatcttttcaggatttatgctctgaagctccttaacatccct 300
 Db |||||
 QY 390 gctctgctctgctgatatcttttcaggatttatgctctgaagctccttaacatccct 449
 Db |||||
 QY 301 gataactaaagatatatttattgtttaacaagacagttgaaggggctagagataca 360
 Db |||||
 QY 450 gataactaaagatatatttattgtttaacaagacagttgaaggggctagagataca 509
 Db |||||
 QY 361 aagagcccaaatccaataggtatttttatttacttgagaacatgcttgggtcaagtca 420
 Db |||||
 QY 510 aagagcccaaatccaataggtatttttatttacttgagaacatgcttgggtcaagtca 569
 Db |||||
 QY 421 tataacatatgctttgagttagaagatagcaatgaaattttccaccagctatacagaacc 480
 Db |||||
 QY 570 tataacatatgctttgagttagaagatagcaatgaaattttccaccagctatacagaacc 629
 Db |||||
 QY 481 ttattttcagttataaacaatggccacaatcagaagtcctatgacatggtgacacct 540
 Db |||||
 QY 630 ttattttcagttataaacaatggccacaatcagaagtcctatgacatggtgacacct 689
 Db |||||
 QY 541 atgagctctattatttgaaggtgatacagtgctctcaggagcttttggatcaggtttta 600
 Db |||||
 QY 690 atgagctctattatttgaaggtgatacagtgctctcaggagcttttggatcaggtttta 749
 Db |||||
 QY 601 gtaaatcgtgtacctgctcacaataaagatttaacaagacgacatgatttggcaaggt 660
 Db |||||
 QY 750 gtaaatcgtgtacctgctcacaataaagatttaacaagacgacatgatttggcaaggt 809
 Db |||||
 QY 661 ttactgaagagacagctcaagctattgagccatatattaccacttttttaacaggtt 720
 Db |||||
 QY 810 ttactgaagagacagctcaagctattgagccatatattaccacttttttaacaggtt 869
 Db |||||
 QY 721 ctgagctcttatttgaaggtgatacagtgctctcaggagcttttggatcaggtttttg 780
 Db |||||
 QY 870 ctgagctcttatttgaaggtgatacagtgctctcaggagcttttggatcaggtttttg 929
 Db |||||
 QY 781 gagctctacaattatgatagctatttgcctcgtctctgttttccccagcttgaattaaa 840
 Db |||||

Db 930 gagctctacaattatgatagctatttgcctcgtctgttttccccagcttgaattaaa 989
 QY |||||
 QY 841 ttaaaagagcaatgataaagaggagcgcctacaagttgttaaaactactgcgcaaaaatggtt 900
 Db |||||
 QY 990 ttaaaagagcaatgataaagaggagcgcctacaagttgttaaaactactgcgcaaaaatggtt 1049
 Db |||||
 QY 901 ggggcaaaaggattcagaattggcttctcaaaaagccactttggcagtgctactgggc 960
 Db |||||
 QY 1050 ggggcaaaaggattcagaattggcttctcaaaaagccactttggcagtgctactgggc 1109
 Db |||||
 QY 961 aggtttaatgatccatgtaccatccgcctgggaatggtgaaaatttctagcactgtt 1020
 Db |||||
 QY 1110 aggtttaatgatccatgtaccatccgcctgggaatggtgaaaatttctagcactgtt 1169
 Db |||||
 QY 1021 ctcatgaaccatccttgatttagcaaaagacttaacagagatctttaaagtgggtcacat 1080
 Db |||||
 QY 1170 ctcatgaaccatccttgatttagcaaaagacttaacagagatctttaaagtgggtcacat 1229
 Db |||||
 QY 1081 gacctgaggaagctattagacatgatttattgtgtcaatgttacagctgctcaaaaag 1140
 Db |||||
 QY 1230 gacctgaggaagctattagacatgatttattgtgtcaatgttacagctgctcaaaaag 1289
 Db |||||
 QY 1141 gatattcttctgtcaatgatccatctacttaatttggtagagagagaacattagacaaa 1200
 Db |||||
 QY 1290 gatattcttctgtcaatgatccatctacttaatttggtagagagagaacattagacaaa 1349
 Db |||||
 QY 1201 cgatggagatgacgcaaaagacatgattgggacttgcctcaaattttataagaaaatgct 1260
 Db |||||
 QY 1350 cgatggagatgacgcaaaagacatgattgggacttgcctcaaattttataagaaaatgct 1409
 Db |||||
 QY 1261 ttacagtcagcagctgggaaaagatgctgcacaaacagatagcatgatacaaaacaaatg 1320
 Db |||||
 QY 1410 ttacagtcagcagctgggaaaagatgctgcacaaacagatagcatgatacaaaacaaatg 1469
 Db |||||
 QY 1321 ctacatatattatcaaaatagatttattgatgactacttggtagacggatcttggct 1380
 Db |||||
 QY 1470 ctacatatattatcaaaatagatttattgatgactacttggtagacggatcttggct 1529
 Db |||||
 QY 1381 caatacatggttctccacaaatttagaaactacagaaacgagatgaaatgcttattacttgg 1440
 Db |||||
 QY 1530 caatacatggttctccacaaatttagaaactacagaaacgagatgaaatgcttattacttgg 1589
 Db |||||
 QY 1441 tatgcacactggattttaaactgctgaaagcattgaaatggtgaaatgctcaaat 1500
 Db |||||
 QY 1590 tatgcacactggattttaaactgctgaaagcattgaaatggtgaaatgctcaaat 1649
 Db |||||
 QY 1501 ctgctccacacatcagtaaaagatttgcctgacttgaattaaagcaccacacacagatgcc 1560
 Db |||||
 QY 1650 ctgctccacacatcagtaaaagatttgcctgacttgaattaaagcaccacacacagatgcc 1709
 Db |||||
 QY 1561 agtgcgaaggcccatattttcaaaagtgatgggttattacaagaaatttacctgactcctggt 1620
 Db |||||
 QY 1710 agtgcgaaggcccatattttcaaaagtgatgggttattacaagaaatttacctgactcctggt 1769
 Db |||||
 QY 1621 aaggctcaggatttcatgaagaaatttcacacaggggtgttagaagatgagagaaataaga 1680
 Db |||||
 QY 1770 aaggctcaggatttcatgaagaaatttcacacaggggtgttagaagatgagagaaataaga 1829
 Db |||||
 QY 1681 aagcagttagaagttacttcttagtccaacatgctcctcgaagcagctgaaagttgtgtg 1740
 Db |||||
 QY 1830 aagcagttagaagttacttcttagtccaacatgctcctcgaagcagctgaaagttgtgtg 1889
 Db |||||
 QY 1741 cgtgaaataaactaaagaagttgggcaaccacacagcctacaaatccttctcgtgaaatg 1800
 Db |||||
 QY 1890 cgtgaaataaactaaagaagttgggcaaccacacagcctacaaatccttctcgtgaaatg 1949
 Db |||||
 QY 1801 atcaagtttctcttgagagagatagcacctgtgcacatagatcacggaatctctcagttgct 1860
 Db |||||
 QY 1950 atcaagtttctcttgagagagatagcacctgtgcacatagatcacggaatctctcagttgct 2009
 Db |||||
 QY 1861 ctatttaacaaagtgaaacaaatcaatagatgaaacagcagatgagagatagaggtgtt 1920
 Db |||||

XX Homo sapiens.
 OS Key Location/Qualifiers
 PH exon 85..130
 FT /tag= a
 FT /number= 1
 FT 131..257
 FT /tag= b
 FT /number= 2
 FT 150..4325
 FT /tag= c
 FT /product= "AS3"
 FT /note= "Androgen shutoff gene 3 protein"
 FT 258..461
 FT /tag= d
 FT /number= 3
 FT 462..548
 FT /tag= e
 FT /number= 4
 FT 549..646
 FT /tag= f
 FT /number= 5
 FT 647..773
 FT /tag= e
 FT /number= 6
 FT 774..938
 FT /tag= f
 FT /number= 7
 FT 939..995
 FT /tag= g
 FT /number= 8
 FT 996..1111
 FT /tag= h
 FT /number= 9
 FT 1112..1206
 FT /tag= i
 FT /number= 10
 FT 1207..1352
 FT /tag= j
 FT /number= 11
 FT 1353..1504
 FT /tag= k
 FT /number= 12
 FT 1505..1618
 FT /tag= l
 FT /number= 13
 FT 1619..1700
 FT /tag= m
 FT /number= 14
 FT 1701..1749
 FT /tag= n
 FT /number= 15
 FT 1750..1889
 FT /tag= o
 FT /number= 16
 FT 1890..2005
 FT /tag= p
 FT /number= 17
 FT 2006..2111
 FT /tag= q
 FT /number= 18
 FT 2112..2272
 FT /tag= r
 FT /number= 19
 FT 2273..2396
 FT /tag= s
 FT /number= 20
 FT 2397..2555
 FT /tag= t
 FT /number= 21
 FT 2556..2624
 FT /tag= u

FT exon /number= 22
 FT 2625..2761
 FT /tag= v
 FT /number= 23
 FT 2672..2885
 FT /tag= w
 FT /number= 24
 FT 2886..3090
 FT /tag= x
 FT /number= 25
 FT 3091..3205
 FT /tag= y
 FT /number= 26
 FT 3206..3338
 FT /tag= z
 FT /number= 27
 FT 3339..3458
 FT /tag= aa
 FT /number= 28
 FT 3459..3521
 FT /tag= ab
 FT /number= 29
 FT 3522..3667
 FT /tag= ac
 FT /number= 30
 FT 3668..3773
 FT /tag= ad
 FT /number= 31
 FT 3774..4213
 FT /tag= ae
 FT /number= 32
 FT 4214..4438
 FT /tag= af
 FT /number= 33
 FT 4314..4408
 FT /tag= ag
 FT /note= "Destabilizing signal"
 FT 4439..5337
 FT /tag= ah
 FT /number= 34
 FT 5312..5317
 FT /tag= ai
 FT /number= 35
 FT 5333..5337
 FT /tag= aj
 FT /note= "Poly adenylation cleavage site"
 XX WO200050454-A1.
 PN 31-AUG-2000.
 PD 24-FEB-2000; 2000WO-US04732.
 PF 24-FEB-1999; 99US-0121461.
 PR (TUFT) TUFTS COLLEGE.
 PA Soto AM, Sonnenschein C, Geck P, Szelei J;
 PI WPI; 2000-565451/52.
 XX P-PSDB; AAY94702.
 DR New human androgen-induced tumor suppressor cDNA sequence termed
 XX 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
 PT diagnosis and treatment of prostate cancer -
 FT Example 4; Fig 6; 152pp; English.
 PS This invention relates to a human androgen-induced tumour suppressor
 XX cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
 CC located on chromosome 13 at position 13q12-13q. AS3 has a role in
 CC inhibiting cell proliferation and use as a marker for the efficient
 CC diagnosis and treatment of prostate cancer. The invention includes AS3
 CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host

Db 66 atggctcattcaagactaggaccatgatggaaaaattacatatccgctgggtcaag 125
QY 61 gaatatcagataaataataaagaggagatggtcagacgattaaagatggttgaaa 120
Db 126 gaaataccagataaataataaagaggagatggtcagacgattaaagatggttgaaa 185
QY 121 actttatgatgatgaccagagactctgaagagaaaaagagctttattttaaacctagct 180
Db 186 actttatgatgatgaccagagactctgaagagaaaaagagctttattttaaacctagct 245
QY 181 ttacatttcttcagatttttttctcaagcatccttggttaaaagatttgcgttacttggtta 240
Db 246 ttacatttcttcagatttttttctcaagcatccttggttaaaagatttgcgttacttggtta 305
QY 241 gctgctgccttgcatttttccagattttatgctcctggaagctctttacacatcccc 300
Db 306 gctgctgccttgcatttttccagattttatgctcctggaagctctttacacatcccc 365
QY 301 gataaactaaagatatattttatgtttaaacagacagttgaaagggtcagaggtataca 360
Db 366 gataaactaaagatatattttatgtttaaacagacagttgaaagggtcagaggtataca 425
QY 361 aagagcccaaatcctaagtggttatttttttatttcttgagaacattgcttgggtcaagtc 420
Db 426 aagagcccaaatcctaagtggttatttttttatttcttgagaacattgcttgggtcaagtc 485
QY 421 tataacatatgctttgatttagaagatagcaatgaaattttccaccagctatacagaacc 480
Db 486 tataacatatgctttgatttagaagatagcaatgaaattttccaccagctatacagaacc 545
QY 481 ttattttcagttataaacaatggcccaatcagaagaatccatagtcacatggttagacctt 540
Db 546 ttattttcagttataaacaatggcccaatcagaagaatccatagtcacatggttagacctt 605
QY 541 atgagctctattttgtgaagtgatcacagtgctctcgaagagcttttgatagcggtttta 600
Db 606 atgagctctattttgtgaagtgatcacagtgctctcgaagagcttttgatagcggtttta 665
QY 601 gtaaatctgtctacgtctacataagaatttaaacagaacacatagtttgcaagagct 660
Db 666 gtaaatctgtctacgtctacataagaatttaaacagaacacatagtttgcaagagct 725
QY 661 ttactgaagaggacagctcaagctatttgagccatatattaccacttttttaacaggtt 720
Db 726 ttactgaagaggacagctcaagctatttgagccatatattaccacttttttaacaggtt 785
QY 721 ctgatgcttgggaaaaacatctacagcatttgcagagcatgcttttgacttaatttg 780
Db 786 ctgatgcttgggaaaaacatctacagcatttgcagagcatgcttttgacttaatttg 845
QY 781 gagctctacaatttgcattgctctctctctctctctctctctctctctctctctctct 840
Db 846 gagctctacaatttgcattgctctctctctctctctctctctctctctctctctctct 905
QY 841 ttaaaagagcaatgataatgaggcgcctacaaagtgtttaaaactactggcaaaaaatgttt 900
Db 906 ttaaaagagcaatgataatgaggcgcctacaaagtgtttaaaactactggcaaaaaatgttt 965
QY 901 ggggcaagagattcagaatttggcttctcaaaaaacagccactttggcagtgctacttgggc 960
Db 966 ggggcaagagattcagaatttggcttctcaaaaaacagccactttggcagtgctacttgggc 1025
QY 961 aggtttaatgatcatcatgtaccatccgcctggaattgtgaaatttgcagccattgt 1020
Db 1026 aggtttaatgatcatcatgtaccatccgcctggaattgtgaaatttgcagccattgt 1085
QY 1021 ctcatgaaccatcctgattttagcaaaagacttaacagagtatcttaaaagtgggtcacat 1080
Db 1086 ctcatgaaccatcctgattttagcaaaagacttaacagagtatcttaaaagtgggtcacat 1145
QY 1081 gaccttgaggaaactattagacatgatgtttattgtgtcatttagctacagctgtctaaaaag 1140
Db 1146 gaccttgaggaaactattagacatgatgtttattgtgtcatttagctacagctgtctaaaaag 1205

QY 1141 gatattcttctggtcactgatcacttacttaattttgtgagagagagaacattagacaaa 1200
Db 1206 gatattcttctggtcactgatcacttacttaattttgtgagagagagaacattagacaaa 1265
QY 1201 cgaatggagagtacgcaagaagccatgatgggacttggcccaattttataagaataatgct 1260
Db 1266 cgaatggagagtacgcaagaagccatgatgggacttggcccaattttataagaataatgct 1325
QY 1261 ttacagtcagcagctggaagaagatgctgcaaaaacagatagctggatcaaaagcaaatg 1320
Db 1326 ttacagtcagcagctggaagaagatgctgcaaaaacagatagctggatcaaaagcaaatg 1385
QY 1321 ctacatatatttatcaaaatagatttgcagacttcttgcagcaggtcttgccttgcct 1380
Db 1386 ctacatatatttatcaaaatagatttgcagacttcttgcagcaggtcttgccttgcct 1445
QY 1381 caatacatggttctcctcacaaatttagaaaactacagaacggtgaaatgcttatattactg 1440
Db 1446 caatacatggttctcctcacaaatttagaaaactacagaacggtgaaatgcttatattactg 1505
QY 1441 tatgcccactggattttaaagctgtgaaaagcattgaatgaaatgtggaatgtcaaaat 1500
Db 1506 tatgcccactggattttaaagctgtgaaaagcattgaatgaaatgtggaatgtcaaaat 1565
QY 1501 ctgctccgacatcaagtaaaagatttgcctgacttgaattaaagcaaccccaaacagatgcc 1560
Db 1566 ctgctccgacatcaagtaaaagatttgcctgacttgaattaaagcaaccccaaacagatgcc 1625
QY 1561 agtgtcaaggccatattttcaaaagtgtgttatcaagaataattacctgactcctggt 1620
Db 1626 agtgtcaaggccatattttcaaaagtgtgttatcaagaataattacctgactcctggt 1685
QY 1621 aaggctcagagatttccatgaagaataatccacaggtgttagaagatgatggaataaaga 1680
Db 1686 aaggctcagagatttccatgaagaataatccacaggtgttagaagatgatggaataaaga 1745
QY 1681 aagcagttagaagatcttctgtagcacaactgctcctgcaagcaggtgaggtgtgtg 1740
Db 1746 aagcagttagaagatcttctgtagcacaactgctcctgcaagcaggtgaggtgtgtg 1805
QY 1741 cgtgaataactaagaagttgggcaaccccaaacagcctacaaatcctttcctggaaatg 1800
Db 1806 cgtgaataactaagaagttgggcaaccccaaacagcctacaaatcctttcctggaaatg 1865
QY 1801 atcaagtttctcttgagagagatagcacctgtgcacatagataccgaaatctacagtgct 1860
Db 1866 atcaagtttctcttgagagagatagcacctgtgcacatagataccgaaatctacagtgct 1925
QY 1861 ctatttaacaagaatgaacaaatcaatagatggaaacagcagatgatgaagatgaggtgtgt 1920
Db 1926 ctatttaacaagaatgaacaaatcaatagatggaaacagcagatgatgaagatgaggtgtgt 1985
QY 1921 ccaactgatcaagccatcagagcaggtcttgaactgcttaaggtactctcattacacat 1980
Db 1986 ccaactgatcaagccatcagagcaggtcttgaactgcttaaggtactctcattacacat 2045
QY 1981 cccatctcatttctctgctgaaaaatttgaatcattactgctgtctgaaaaatggat 2040
Db 2046 cccatctcatttctctgctgaaaaatttgaatcattactgctgtctgaaaaatggat 2105
QY 2041 gatgaaaaagtagcagaagctgcactacaaattttcaaaaacacaggaagcaaaatgaa 2100
Db 2106 gatgaaaaagtagcagaagctgcactacaaattttcaaaaacacaggaagcaaaatgaa 2165
QY 2101 gaggattttccacacatcagatcagccttgccttgcctgttttatcatcacaaatctaaaaa 2160
Db 2166 gaggattttccacacatcagatcagccttgccttgcctgttttatcatcacaaatctaaaaa 2225
QY 2161 gggccccccgctcaagccaaatagtcattcattgtatccatcgatattttctagtaa 2220
Db 2226 gggccccccgctcaagccaaatagtcattcattgtatccatcgatattttctagtaa 2285


```
FT exon 465..562
FT /*tag= f
FT /number= 5
FT exon 563..689
FT /*tag= e
FT /number= 6
FT exon 690..854
FT /*tag= f
FT /number= 7
FT exon 855..911
FT /*tag= g
FT /number= 8
FT exon 912..1027
FT /*tag= h
FT /number= 9
FT exon 1028..1122
FT /*tag= i
FT /number= 10
FT exon 1123..1268
FT /*tag= j
FT /number= 11
FT exon 1269..1420
FT /*tag= k
FT /number= 12
FT exon 1421..1534
FT /*tag= l
FT /number= 13
FT exon 1535..1616
FT /*tag= m
FT /number= 14
FT exon 1617..1665
FT /*tag= n
FT /number= 15
FT exon 1666..1805
FT /*tag= o
FT /number= 16
FT exon 1806..1921
FT /*tag= p
FT /number= 17
FT exon 1922..2027
FT /*tag= q
FT /number= 18
FT exon 2028..2188
FT /*tag= r
FT /number= 19
FT exon 2189..2312
FT /*tag= s
FT /number= 20
FT exon 2313..2471
FT /*tag= t
FT /number= 21
FT exon 2472..2540
FT /*tag= u
FT /number= 22
FT exon 2541..2677
FT /*tag= v
FT /number= 23
FT exon 2678..2801
FT /*tag= w
FT /number= 24
FT exon 2802..3006
FT /*tag= x
FT /number= 25
FT exon 3007..3121
FT /*tag= y
FT /number= 26
FT exon 3122..3254
FT /*tag= z
FT /number= 27
FT exon 3255..3374
FT /*tag= aa
FT /number= 28
FT 3375..3437
```

```
FT /*tag= ab
FT /number= 29
FT exon 3438..3583
FT /*tag= ac
FT /number= 30
FT exon 3584..3689
FT /*tag= ad
FT /number= 31
FT exon 3690..4129
FT /*tag= ae
FT /number= 32
FT exon 4130..4354
FT /*tag= af
FT /number= 33
FT misc_signal 4320..4324
FT /*tag= ag
FT /note= "Destabilizing signal"
FT exon 4355..5253
FT /*tag= ah
FT /number= 34
FT polyA_signal 5228..5233
FT /*tag= ai
FT misc_feature 5249..5253
FT /*tag= aj
FT /note= "Poly adenylation cleavage site".
FT XX
PN WO200050454-A1.
XX 31-AUG-2000.
XX 24-FEB-2000; 2000WO-US04732.
XX 24-FEB-1999; 99US-0121461.
XX (TUFT ) TUFTS COLLEGE.
XX Soto AM, Sonnenschein C, Geck P, Szelei J;
XX WPI; 2000-565451/52.
XX P-PSDB; AAY94702.
XX New human androgen-induced tumor suppressor cDNA sequence termed
XX 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
XX diagnosis and treatment of prostate cancer -
XX Claim 1; Fig 1; 152pp; English.
XX This invention relates to a human androgen-induced tumour suppressor cDNA
XX sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
XX on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
XX proliferation and use as a marker for the efficient diagnosis and
XX treatment of prostate cancer. The invention includes AS3 cDNA and protein
XX sequences, a vector comprising the cDNA sequence, a host cell transfected
XX with the expression vector, and a method for producing an AS3 polypeptide
XX comprising culturing the transfected cells. AS3 has cytostatic activity,
XX and acts to suppress cell proliferation. The AS3 gene is useful as a
XX marker for the efficient diagnosis and treatment of prostate cancer. The
XX AS3 nucleic acid molecule can be used as a source of antisense agents. For
XX sequence specific modulation of gene expression. The AS3 protein may be
XX used in the treatment of disorders caused by aberrant modification or
XX mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
XX or aberrant post-translational modification of the AS3 protein. This
XX sequence represents the human AS3 cDNA sequence.
XX Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;
```

Query Match 100.0%; Score 4173; DB 21; Length 5271;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctattcaaacactagaccatgatggaaaaattacatatccgctgggtcaag 60
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 19:52:23 ; Search time 520.53 seconds
(without alignments)
5033.780 Million cell updates/sec

Title: US-09-512-581-3
Perfect score: 4173
Sequence: 1 atggctattcaagactag.....cacaacccaataaataatgtg 4173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4173	100.0	5271	21	AAA28051 Human androgen shu
2	4173	100.0	5355	21	AAA28052 Human androgen shu
3	1112.8	26.7	3957	21	AAA47423 Sequence encoding
4	411.4	9.9	439	20	AAV87629 EST clone DY17. H
5	348	8.3	738	20	AAZ15259 Human gene express
6	298.8	7.2	351	21	AAZ42861 Human 5' EST isola
7	292	7.0	295	21	AAC02817 Human secreted pro
8	253.8	6.1	1217	21	AAF18338 Lung cancer associ
9	203.6	4.9	530	21	AAZ80598 Human colon cancer
10	133.2	3.2	936	22	AAF58252 Oligonucleotide D1
11	133.2	3.2	936	22	AAF58254 Oligonucleotide D1

c	12	133.2	3.2	936	22	AAF58257	Oligonucleotide D1
c	13	133.2	3.2	936	22	AAF58259	Oligonucleotide D2
c	14	133.2	3.2	936	22	AAF58262	Oligonucleotide D2
c	15	133.2	3.2	938	22	AAF58255	Oligonucleotide D1
c	16	132.4	3.2	936	22	AAF58252	Oligonucleotide D1
c	17	132.4	3.2	936	22	AAF58254	Oligonucleotide D1
c	18	132.4	3.2	936	22	AAF58257	Oligonucleotide D1
c	19	132.4	3.2	936	22	AAF58259	Oligonucleotide D2
c	20	132.4	3.2	936	22	AAF58262	Oligonucleotide D2
c	21	132.4	3.2	938	22	AAF58255	Oligonucleotide D1
c	22	99	2.4	161	21	AAC06611	Human secreted pro
c	23	59.2	1.4	6644	20	AAZ33181	Base sequence of t
c	24	59.2	1.4	7372	20	AAZ33182	Base sequence of t
c	25	59.2	1.4	7797	20	AAZ33180	Cowpox virus bsr f
c	26	59.2	1.4	7996	20	AAZ33184	Base sequence of t
c	27	57.6	1.4	4000	18	AAZ191902	Mannose-1-phosphat
c	28	54.8	1.3	244	22	AAF58238	Oligonucleotide D1
c	29	54.6	1.3	6755	19	AAV21511	staphylococcal bac
c	30	54.4	1.3	244	22	AAF58238	Oligonucleotide D1
c	31	53	1.3	5059	20	AAZ84332	stealth virus nucl
c	32	51.6	1.2	3399	17	AAZ05868	Chicken leucocytoz
c	33	51.4	1.2	72604	20	AAZ10752	Genomic sequence o
c	34	48.6	1.2	396	22	AAF94862	Human ovarian can
c	35	48.2	1.2	1236	21	AAZ02163	Human colon cancer
c	36	48.2	1.2	10732	21	AAZ10594	Gene encoding a su
c	37	48.2	1.2	49999	20	AAZ23895	Murine LOBO homolo
c	38	47.6	1.1	6741	21	AAZ10595	Gene encoding a su
c	39	47.2	1.1	1686	16	AAZ07587	DNA encoding leuco
c	40	45.8	1.1	49999	20	AAZ23891	Murine LOBO homolo
c	41	45.8	1.1	49999	20	AAZ23896	Murine LOBO homolo
c	42	45.6	1.1	979	20	AAZ61816	B. burgdorferi ant
c	43	45.6	1.1	1083	20	AAZ61815	B. burgdorferi ant
c	44	45.6	1.1	2487	20	AAZ20298	Borrelia burgdorfe
c	45	45.2	1.1	43576	21	AAZ68609	Bacteriophage 96 c

ALIGNMENTS

RESULT 1
AAA28051
ID AAA28051 standard; cDNA; 5271 BP.
XX
AC AAA28051;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX
DE Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW diagnosis; treatment; cytostatic; human; ss.
XX
OS Homo sapiens.
XX
FH Key
FT exon
FT Location/Qualifiers
FT 1..46
FT /*tag= a
FT /number= 1
FT 47..173
FT /*tag= b
FT /number= 2
FT 66..4241
FT /*tag= c
FT /product= "AS3"
FT /note= "Androgen shutoff gene 3 protein, the CDS is specifically claimed as SEQ ID #3"
FT 174..377
FT /*tag= d
FT /number= 3
FT 378..464
FT /*tag= e
FT /number= 4

Wed Sep 26 11:24:08 2001

us-09-512-581-3.rst

Page 14

REFERENCE 1 (bases 1 to 888)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: L10CM1558 row: P column: 11
High quality sequence stop: 688.

FEATURES

Source

```
1. 888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4714234"
/clone_lib="NIH_MGC_79"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Placenta; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATCTAGAGCGCCGAGCGCCGACATG-dT(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
```

BASE COUNT 347 a 163 c 223 g 149 t
ORIGIN

Query Match 11.2%; Score 468; DB 155; Length 888;
Best Local Similarity 99.8%; Pred. No. 3.4e-103;
Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 3693 ggaatgaactgaactaagtgtgtacaggaacagaacaaacctaaggcagtcggaagtcgga 3752
Dc 1 GGATGACTTGACTAGTGTGTACAGGACAGAAACCTAAAGGCACTCAGCGAATCGGAA 60
QY 3753 aagagggcattcggtctcaagaattgtatgaacagcagctggcctgaagaaagagctcaa 3812
Dc 61 AAGAGGCCATCGGCTTCAGATCTGATGAACAGCAGTGGCTTGAGAAAGAGCGCTCAA 120
QY 3813 agaagatatattagaanaatgaagatgaacagaatagtcgcgcaaaaaaagggttaaaagg 3872
Dc 121 AGAAGATATATTAGAAATGAAATGAAACAGATAGTCCGCC-AAAAAGGTTAAAGAGG 179
QY 3873 ccgaaccacaaacctcttggctggagtgacacaaagaagaagccaacatgaaacttc 3932
Dc 180 CCGACACCAACAAACCTTGTGTGGAGGTACACCAAAAGAGAGCCACACATGAAATCTTC 239
QY 3933 taataaagaagaacaaaaaatacttggaactccagcaccagaagaagaagaagaaga 3992
Dc 240 TAAAAAAGGAAGCAAAAAAATCTGGACTCCAGCACAGAGGAGGAAGGAAGA 299
QY 3993 aagacaagtgagaatacgaacagaagaagtcacaaagcaaacagacccgagttcaaggag 4052
Dc 300 AAGACAAAGTGGAATACGGAACAGAACTCCAAAAGCAAAACGACCGAGTCTCAAGGAG 359
QY 4053 agcacaagcagaagaagaatctctgaaattagtgaaattgaattcacacagttccacacc 4112
Dc 360 AGCACAGCAGAGAGAGATCTCTGAAATCTAGTCAATTGAAATCCACACAGTCCACACC 419
QY 4113 acagaagaagcaggaagaacacatcaaaaaagccatcacatcaacaacaaaaaaatgt 4172
Dc 420 ACAGAAAGAGAGAGAGAGACCATCAAAAAAGCCATCACCATCACAACCAAAAAAATGT 479
```

Search completed: September 25, 2001, 16:19:35
Job time: 6023 sec

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.
 BDNALDO, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
 www.bio.linnl.gov/bdnp/image/image.html
 Insert length: 1048 Std Error: 0.00
 Seq primer: -40UP from Gldco
 High quality sequence stop: 455.

FEATURES

Source
 1. .480
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2239331"
 /clone_lib="NCI CGAP GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bernaldo."

BASE COUNT

82 a 123 c 84 g 191 t

Query Match 11.5%; Score 478.4; DB 23; Length 480;
 Best Local Similarity 99.8%; Pred. No. 8.4e-106;
 Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3460 acctgaagaatgcgaagcagctcaatccaaagctctctggaagaataaaggagg 3519
 |||||||
 480 ACTGTAACCAATGCAGAGCAGCTCAATCAACCTCTCGGAAGATAAAGGGAGG 421
 |||||||
 3520 cttaatacttgaatgatcatcagtgaaatgaatttcacaaatgtctacccttg 3579
 |||||||
 420 CTGATAGCTTCTGAAATGATCAGTGAATAATGAAGATACACAATGCTTCACCTTTG 361
 |||||||
 3580 ccggggaaaaaagtgcagaagagacactctgactgtaaggtctcgaattgagaag 3639
 |||||||
 360 CCGGGGAAAAAAGTGCAGACAGACACTCTGATCTTGTAAGCTCTGAATTGAGAA 301
 |||||||
 3640 cctgaagcaggaaaaaaacgcccgtcaacagacagagagagaatagatgagac 3699
 |||||||
 300 CCTGAGAGCAGGAAAAAAGCGCCCTCACAGAACAGAGAGAGAAATTAAGTATGATGAC 241
 |||||||
 3700 ttgactaagtgtgataagaagaacaaacaaagcagtcagcagagtcggaagaagagc 3759
 |||||||
 240 TTGACTAAGTGTGATACAGACAGAAACCTTAAGGACATCAGCAAGTCGCAAAAGAGGC 181
 |||||||
 3760 catacagctcagaatcgtatgaacagcagtgctgaggaagaagagctcaagaagat 3819
 |||||||
 180 CATACGCTTCAAGATCTGATGATGACAGCATGTGCTGAGGAAAAAGAGCTCAAAAGAT 121
 |||||||
 3820 atattagaanaatgaagatgaacagaatagtcgcgcaaaaaagggtlaaagaagccgacaa 3879
 |||||||
 120 ATATTAGAANAATGAAGATGACAGCATAGTCCGCAAAAAAGGTTAAAGAGAGCCGACCA 61
 |||||||
 3880 ccaaaacctctgtgtgaggtgacaccaaagaagagccaaacatgaataactcttaaaaa 3939
 |||||||
 60 CCAAAACCTCTGTGTGAGGTACACCAAAAGAGCCAAAAATGAAAACTTCTAAAAA 1

RESULT 13
 LOCUS BF471938 608 bp mRNA EST 04-DEC-2000
 DEFINITION UI-M-BH3-avr-g-03-0-UI.r1 NIH_BMAP_M.S4 Mus musculus cDNA clone
 UI-M-BH3-avr-g-03-0-UI 5', mRNA sequence.

ACCESSION BF471938
 VERSION BF471938.1 GI:11541121
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 608)
 REFERENCE Aurhors
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 9704447
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestrail.nih.gov

FEATURES

Source

Location/Qualifiers
 1. .608
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH3-avr-g-03-0-UI"
 /clone_lib="NIH_BMAP_M.S4"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M.S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M.S4, NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, NIH_BMAP_M.S3.1, NIH_BMAP_M.S2, NIH_BMAP_M.S1. The subtracted library (NIH_BMAP_M.S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, and NIH_BMAP_M.S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M.S3.3, NIH_BMAP_M.S3.2, and NIH_BMAP_M.S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, into DH10B bacteria (Life Technologies) and electropotated into NIH_BMAP_M.S4 library. This procedure has been previously described (Bernaldo, Lennon and Soares, Genome Research 6:791-806, 1996)."

BASE COUNT 193 a 113 c 128 g 174 t

Query Match 11.4%; Score 477.4; DB 149; Length 608;
 Best Local Similarity 90.9%; Pred. No. 1.6e-105;
 Matches 508; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

LOCUS 1 atggtcatcctaagaactagacccaatgataatcaatcgcgctgggtcaag 60
 |||||||
 50 ATGCTCATTTCAAAAGACCAAGATGGGAAATTAATTACTTACCTTCGTGACTCAAG 109

OY	3109	cacgctttatcaagaaggtggtaaaaaattaaacaacaaaagaatcccaaggcca	3168
Dc	181		
OY	3169	gatga tgcnaaaatgaatgaanaactgtcacactgtgtgtgcatgttgcca tgaata ctc	3228
Dc	241	GATGATACAAAATAAATGAATCAAAATTGTTCACCGTGTGTGATGTGCCATGAACATCATC	300
OY	3229	atgtcaaagagtactacatacagtttggaaattctcctaagaaccggtactaccaactgtg	3288
Dc	301	ATGTCAAAGAGACACACGCACGCTGGAGTGTCTCCTAAGACCCCGTGCTGCACGCTCGG	360
OY	3289	tctctcaactcaaccggaacaagaattcag taaacacccaanaattatctgctcctgaaty	3348
Dc	361	TTTTTTCACCACCCCGCACAGAATTTTACTAACACCAAAAATTTACTGCTCTCAGAAAA TG	420
OY	3349	aatacatcttctactcctcygnaaaccttaaacaaccaaatlytcttagagcty tlaaca ag	3408
Dc	421	AATATATTTTTCACCTCCGGAATAACCTTAAMAACGCCAATGTCTCGGAGCACTATTA TAG	480
OY	3409	ccacttcatcagaggcgcaagaactctgacccaatatc tcaagaaatgnaa actgaa ac	3468
Dc	481	CCACTTTATTCAGCAGGCAACACAGTCTCAGACAATTCATCAAGATGGAAACTGTGAC	540
OY	3469	aatgaa -gcagcagctcaaatccaa gctctcctcygnaaataaagggagagcttga tag	3527
Dc	541	AACGCAANGCACGACTCCAAACCCCAAGCTCTCTGAGAGGATCAAGGAGGCTTGATTA	600
OY	3528	ttctgaatgatactacagtgaataatgaagat	3558
Dc	601	GCTCTGAATGGTCACTGAAACATGAAGAT	631
 RESULT 11 BB551003			
Locus	BE551003	535 bp	mRNA EST 10-AUG-2000
DEFINITION	7b67bd03.x1 NCI-CGAP_Lu24 Homo sapiens CDNA clone IMAGE:3233261 3'		
ACCESSION	BE551003		mRNA sequence.
VERSION	BE551003		
KEYWORDS	BE551003.1 GI:9792695		EST.
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN, send email to:		
	info@image.lnl.gov		
	Seq primer: -40up from Glibco		
	High quality sequence stop: 475.		
FEATURES	Location/Qualifiers		
source	1..535		

Query Match	11.6%	Score 482.6;	DB 137;	Length 535;
Best Local Similarity	96.6%;	Pred. No. 8.2e-107;		
Matches 514;	Conservative 0;	Mismatches 15;	Indels 3;	Gaps 2;
Oy 2651	aatcatccattagacacatatcagctctgctcctatgctctcgaagatgaatgtcatca	2750		
Db 7	AAACATCATATTAGACA--TTCACTATATGTCATTAACGATGATGCTATATC	64		
Oy 2751	aglaagacaagtglttggccagaacatcacaagaagcccttccglttacagcttccact	2810		
Db 65	AAGTAGACAAAGTGTTGGCCAGAAACTT-ACAAAAGCCCTTCCGGTTAGCGCTTCCATT	123		
Oy 2811	tgaatataatgacatctctgccccttctgcaaaagaatccgttaagaagagaagaagctca	2870		
Db 124	TTAGTATATGCAATCTGTGCCCTTTGTGCAAAAAGATCTTTAAAGGAGAAAGAGCTCA	183		
Oy 2871	tgctaggaacagtttggtagaaataataatgtaaggcggaatctcgaagcagcatgc	2930		
Db 184	TGCTAGGCAATGTTTGGTGCAAAAATATTAATGTAAGCGGAGATCTTGAAACACATGC	243		
Oy 2931	agctgttagtgaaaaaatattgtctctctcaaccagatcagltgttccatatcaaatca	2990		
Db 244	AGCTGTATGTAATAAATATTGTCTCTTCCACGAGATGTTGTCATATACAAATTC	303		
Oy 2991	cccttggcaatagaccagaattatgtaaaagtcagatatltgaacaacttaagaatgt	3050		
Db 304	CCTTTGGCACATGACCCAGATATGTCAAGATACAGATATGACAACTTAAGATGT	363		
Oy 3051	taaaagaatgctcttggctgtcttctggaataataaaatggtcctaaagaatgaataacagtca	3110		
Db 364	TAAAGAATGCTCTTGTGTTGTCTGGAATATTTAATGCTGAAATGAAATTAACAGTCA	423		
Oy 3111	cgctttatcgaagaatggtgtagaataataataaacaacaaagaatggcccaagaccga	3170		
Db 424	CGCTTTATCGAAAGATGTCGAAATTTTAAACAAACAAAGATGCCCAAGGACGACA	483		
Oy 3171	tgaatcaaaatgaatgaanaaactgtacacgtgtgtgaatgttgcacgaat	3222		
Db 484	TGATCAAAAATGAAATGAAAAACTGTACACTGTGTCTATGTGGCATGAT	535		

BASE COUNT 183 a 90 c 108 g 153 t 1 others

ORIGIN

library NCI CGAP. L45 was prepared, and ss circles were made in vitro. Following BAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 141920-1417991 and 1520904-1522429). Subtraction by Bento Soares and M. Fatima Bernaldo.

AV685306
 LOCUS AV685306 608 bp mRNA EST 25-SEP-2000
 DEFINITION AV685306 GKC Homo sapiens cDNA clone GKAT08 5', mRNA sequence.
 ACCESSION AV685306
 VERSION AV685306.1 GI:10287169
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 608)
 AUTHORS Wu, T., Qian, B., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z., and Han, Z.
 TITLE Homo sapiens cDNA GK- clones
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919 (ex. 45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source 1. 608
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GKAT08"
 /clone_id="GKC"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 181 a 113 c 125 g 189 t
 ORIGIN
 Query Match 12.1%; Score 503.8; DB 32; Length 608;
 Best Local Similarity 96.2%; Pred. No. 5.7e-112;
 Matches 528; Conservative 0; Mismatches 17; Indels 4; Gaps 1;
 Oy 385 ttttattacttgaagaattgtgtgtgctacatgataataataatgctttgaattgaa 444
 2 ttttattacttgaagaattgtgtgtgctacatgataataataatgctttgaattgaa 61
 Oy 445 gatagaatgaatttaccacagctatagagaacctatttccagttataaacaatgac 504
 62 gatagcaatgaatttaccacagctatagagaacctatttccagttataaacaatgac 121
 Oy 505 cacaatcagaatgctacatgacatgtagaagctttagagctctcattatgtgaagtc 564
 122 cacaatcagaatgctacatgacatgtagaagctttagagctctcattatgtgaagtc 181
 Oy 565 gataagtgctcagaagctttagagctttagagctttagagctttagagctttagag 624
 182 gatagcaatgctcagaagctttagagctttagagctttagagctttagagctttagag 241
 Oy 625 aatttaacaagaatgctacatgatttgcgaagctttagagctttagagctttagag 684
 242 aatttaacaagaatgctacatgatttgcgaagctttagagctttagagctttagag 301
 Oy 685 attgagcattatatacaacttttttaatacaggttctgtagcttgggaacaacatc 744
 302 attgagcattatatacaacttttttaatacaggttctgtagcttgggaacaacatc 361
 Oy 745 agcgatttgaagcagtgctttagagctttagagctttagagctttagagctttagag 804
 362 agcgatttgaagcagtgctttagagctttagagctttagagctttagagctttagag 421
 Oy 805 ttgcgtcctctgttttaccacagcttgaatttaataagaagaatgataatgagag 864

|||||
 Db 422 TTGCTGCTCTCTGCTTACCCAGCTTGAATTAACATAAGAGCAATGATGAGAG 481
 Oy 865 cgcctacaagtgtttaaactactgagcaaaaatgtttggcgcaagaatcagaattgct 924
 Db 482 cgcctacaagtgtttaaactactgagcaaaaatgtttggcgcaagaatcagaattgct 537
 Oy 925 tctcaaac 933
 Db 538 TCTCCATC 546
 RESULT 10
 LOCUS BE532986 603 bp mRNA EST 09-AUG-2000
 DEFINITION 601235013P1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3598922 5', mRNA sequence.
 ACCESSION BE532986
 VERSION BE532986
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 603)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-rt@mail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMN at:
 http://image.llnl.gov
 Plate: LLM8780 row: 1 column: 03
 High quality sequence stop: 607.
 FEATURES
 source 1. 683
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3598922"
 /clone_id="NCI_CGAP_Mam6"
 /sex="Female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
 BASE COUNT 235 a 146 c 138 g 163 t
 ORIGIN
 Query Match 11.9%; Score 495; DB 136; Length 683;
 Best Local Similarity 88.6%; Pred. No. 8.3e-110;
 Matches 559; Conservative 0; Mismatches 70; Indels 2; Gaps 2;
 Oy 2930 cagctgttagtgaataatatt-tgtccttctacagaagtagtgttccatacaatt 2988
 Db 1 CAGCTGTAGTGAATAAATATATGTCTCTTACCAAGATATGTTCCATATACAAAT 60
 Oy 2989 caaccttggcacaagcacaatattatgcaagaagtagatggaacaactaaagat 3048
 Db 61 CACCTTTGGCACAAGCACCACATATATGCAAGAAGTAGAGATATGACCACTTAAAGT 120
 Oy 3049 gttaagaatgtcttctgtgttcttctggaatattaatgcttaaaatgaataacagt 3108
 Db 121 GTGAAGAATGCTTGTGTTGTCTGAGATATGATGATGCTAAATAAAGCAACAGC 180

Matches 572; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 2687 atgaatcatcattacatacaatcagctatgctatgcatatcatcaacgaatgct 2746
 1 ATTAATCATCATATAGAACATATCAGTATGTCATTAAGTATCAACGATGATGCT 60
 OY 2747 atcaagtaagaagtgtttgcccagaacattcaaaaagcccttccglttacggctc 2806
 61 ATCAAGTAAGACAAAGTGTGGCCCAAGAACTTCACAAAGCCCTTCCCGTTACGGCTTC 120
 OY 2807 cacttagatagatggaactctgccccttctgtccaaaagatccttgaagaaggaagag 2866
 121 CACTTAGATATAGCAACTCTGCCCCTTGTGCAAAAAGTCTGTAAAGAGAGAGAG 180
 OY 2867 ctcatactgagcaatgtttgtgtaaaaataataatgtaagcgaggatcttaagcagc 2926
 181 CTCATCTCCTANCGCAATGTTGTGTAATAATTAATGTAAGCGCGGATCTCAACAGC 240
 OY 2927 atgcagctgttagtgaataaattatgtctcttaccagagtaagtgttccatatacaa 2986
 241 ATGCAGCTGTGTGTAATAATTAATGTCCTTACACAGATGATGTTCCATATACAA 300
 OY 2987 ttcaacctttgtgacagtcagccagatgatgtaagtaagtaagtaagtaagtaag 3046
 301 TTCACCTTTGTGACAGTACCCAGATTAATGTAAGTACAGATTAATGAACAACCTTAAG 360
 OY 3047 atgttaagaatgcttctgtgtctgtgtaaaataataatgtaaaataataa 3106
 361 ATGTTAAAGAAATGCTTTGTGTTGTTCTGTAATTAATGCTTAATAATGAATAATAC 420
 OY 3107 gtcaagctttatcagaagaatgtgtaaaataataatacaaaaagaatgtcccaagac 3166
 421 GTCAAGCTTTATCAAGAAAGATGTGTAATAATTAATGAACAACAAGATGCCAATGAC 480
 OY 3167 cagatgatgcaaaaatgtaaaataataataataataataataataataataata 3226
 481 CAGATGATGCAAGAAAGATGAATAAATCTGACCTGTGTGATGTTGCCAATAATAC 540
 OY 3227 tcatgtcaagaagtactacatac-aagtgtgaatcc 3263
 541 TCATGTCAAGAAGTACTACATACNACTTGTGAATCTCC 578
 DE 541 TCATGTCAAGAAGTACTACATACNACTTGTGAATCTCC 578

RESULT 8

BF142489 944 bp mRNA EST 24-OCT-2000
 LOCUS 601788573f1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4016155 5',
 DEFINITION mRNA sequence.
 ACCESSION BF142489
 VERSION BF142489.1 GI:10981529
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 944)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM9263 row: a column: 20
 High quality sequence stop: 620.
 Location/Qualifiers

FEATURES
 source 1..944
 Location/Qualifiers
 /organism="Mus musculus"

/strain="CZECH II"
 /db_xref="taxon:10090"
 /clone="IMAGE:4016155"
 /clone_1ib="NCI_CGAP_Lu30"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /note="Organ: Lung; Vector: PCMV-SPO6; Site:1: NotI;
 Site:2: SalI; Transgenic model MNT-1, expression driven by
 MMTV-ltr enhancer; Cloned unidirectionally. Primer: Oligo
 dT. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 277 a 221 c 218 g 228 t

ORIGIN

Query Match 12.8%; Score 532.8; DB 145; Length 944;
 Best Local Similarity 86.7%; Pred. No. 5.5e-119;

Matches 633; Conservative 0; Mismatches 92; Indels 5; Gaps 4;

OY 2492 tgaatgtcagatgctacttggaataaataatacaacgaataacgaactcact 2551
 1 TGATGTTGATGCTACTTGGATGAATGAATATCAGATAGTCAGGAACCTCCACC 60
 OY 2552 taagattgtaacaacaataatgcatagatgagagacttgacagacagggaaacta 2611
 61 TCAGATTACTAACACAGATCTCATGATGATGGGATTTGACAGAAACAAAGGAAATTA 120
 OY 2612 gtaaacagatagatgactctggaacttgctgtggagtgctatgtgaagctggac 2671
 121 GTAACCGATATGTCACGCTGAGACTGTCTGAGAGTGTCTATTTGTAACCTGGCAC 180
 OY 2672 aagaaccctgttaccatgaatacatcatagaaacaataacagcatgtcatagta 2731
 181 AGGAGCCCTGTTACACAGATGATTAACATGAGCAGTACACAGCTGTGGCATTAAGCCA 240
 OY 2732 tcaagatgaatgctatcaagtaagaagaatgttgcagaagaacttcagaagcctt 2791
 241 TCATATGATGATGTTATCAAGTCAGGACAGTGTCTCTAGAAACTTCACAAAGGCTTT 300
 OY 2792 cccgttaccgcttccacttgaatgatatgcaactcgttcccttgtgcagaagaactc 2851
 301 CCCGCTTACGGCTTCCCTTTGAGTACATGAGCCATCTGTCTTTTGTGCCAAAGACCTG 360
 OY 2852 taaagagagaagaagctcactgtctaggaatgttgggtgaanaaatalaatagtlaagcgagg 2911
 361 TGAAGAAGAGGCGAGCCATGTAGACAGTGTGGGAAGAACATCACTGTGAGAGGG 420
 OY 2912 agtatctgaagcagcagctgcaagctgttaagtaaaatattgtctcttaccagagatag 2971
 421 AGTACCTGAAGCAGCATGACGTAGTGAATAATTAATTTGTTCTTACACAGAGTATG 480
 OY 2972 ttgtccatatacaattcacttctgtgcacatgaccagatgacatgtcacaagtacagata 3031
 481 TGGTTCATATCAATTCACCTTTGGGACATGACCCAGATTAATGTGCAACGTACAGATA 540
 OY 3032 tgaacaacttaagaatgttaagaatgttctgtgttctgtcgaataataatagtgcta 3091
 541 TTGAACAACCTTAAGAATGTGACAAAGATGCTTGG--TTGTTGTGAGAAATTAATGATG 599
 OY 3092 aaaaataaatacaatcagcttcttataagaagaatgtgaagaataataaacaaca 3151
 600 AAAATATAAACAACAG-CATGCTTTATATCAGAAATATGTAAGAAATTAATTAACAGACCA 658
 OY 3152 aagaatcccaagaacagatgcatcaaaaataatgaataaagctgacactgtgtgagatg 3211
 659 AGATTCTCTCAGGAGCAGATGAT-CACCACTGACTCTCACAACTGT--CCCGTGTGTGATG 715
 OY 3212 ttgcacatgaa 3221
 716 TTGCCATGGA 725

RESULT 9

QY 543 gaccctattatttgcgaaggatgatacagtgctcaggaagctttggatacgglttagt 602
 |||||||
 DB 180 GACCTATATATTTGTAAGGTGATGATGCTCAGGACCTTTGGATACGGTTTAGT 239
 |||||||
 QY 603 aactctgtaaccgtctcataagaatttaacaagcaagcatgatatttgcagaagctt 662
 |||||||
 DB 240 AAATCTGATACCTGCTCATGAATTTAAACAAGCAAGCATATGATTTGGCAAGGCTTT 299
 |||||||
 QY 663 actgaagagagacagctcaagcttatgacatataaccacttttatacaggtct 722
 |||||||
 DB 300 ACTGAAGAGACAGCTCAAGCTATTTGACCATATTTTATACCAATTTTATACAGGTTCT 359
 |||||||
 QY 723 gatgcttgggaaacaacatcatcagcagatttgcagagatgctc-tttgacttaatttgg 781
 |||||||
 DB 360 GATGCTTGGGAAAACATCATCAGCATTTGTCAGAGACATGCTTTTGGACTTAATTTGG 419
 |||||||
 QY 782 agctctacacatattgtagtcaatttgcctcctccttcttaccacgcg-ttgaatttaa 840
 |||||||
 DB 420 AGCTCTACATATTTGATGATGATTTGCTGCTCTGTTTAAACCCAGCTTTGAATTTAA 479
 |||||||
 QY 841 ttaagaagcaatgataatgagagcgacacagcttgcataactgacagcaaatgctt 900
 |||||||
 DB 480 TTAAAGAGCATATATGATGAGGAGCCCTTACAGTTCTTAACTACTGCAAAAATGTTT 539
 |||||||
 QY 901 ggggcaagagatcaggaattgctctc-aaaaaagcaacttggcagtgctac-ttgg 958
 |||||||
 DB 540 GGGGCAAGGATTCAGAAATTTGGCTTCTCAAAAACAGCCACTTTGGCAGTGCTACTTTGG 599
 |||||||
 QY 959 gaaggttaataatgatacctatgatac-ccgcctggaatgctga-aatttgcctgcca 1016
 |||||||
 DB 600 GCGAGTTTAAATGATTCACATGACCAATTCGCTGGAATGTGACCAATTTCTAGCCA 659
 |||||||
 QY 1017 ttgtccatgaacacatccctgatttgc-aaaagacttaacagatatactt-aaagtgaag 1074
 |||||||
 DB 660 TTGCTCATGAGCAACCATCGATTACCAAAAAGATTTAAGAGATATCTTAAAGTGAGG 719
 |||||||
 QY 1075 tcaatgacc 1084
 |||||
 DB 720 TCACCTTGAAAC 729
 |||||
 RESULT 5
 BE504550 596 bp mRNA EST 04-AUG-2000
 LOCUS hz58a05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3212144 3'
 DEFINITION similar to TR:094237 O94237 HYPOHETICAL 45.2 KD PROTEIN ;, mRNA
 sequence.
 ACCESSION BE504550
 VERSION BE504550.1 GI:9706958
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 596)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 458.
 FEATURES
 source Location/Qualifiers
 1..596

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3212144"
 /clone_lib="NCI_CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Plasmid DNA from the normalized
 library NCI_CGAP_Lu5 was prepared, and ss circles were
 made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (clones
 1414920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 211 a 103 c 117 g 165 t
 ORIGIN

Query Match 14.3%; Score 596; DB 136; Length 596;
 Best Local Similarity 100.0%; Pred. No. 1.9e-134;
 Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2687 atgaatcatcaccatagaacatatacgtatgctcattagctatcaacgatgatgct 2746
 |||||||
 DB 1 ATGAATCATCATCATTAAGAAACAATATCAGCTATGCTAGCTATCAACGATGAATGCT 60
 |||||||
 QY 2747 atcaatgaagaagaagtgctggcccaagaacttcaagaagccttccgtttacgcttc 2806
 |||||||
 DB 61 ATCAAGTAAACAAAGGTTTGCCAGAACTTCACAAAGGCTTTCCGTTTACGGCTTC 120
 |||||||
 QY 2807 cacttgaglatatgcaatcgtgccccttctgcaaaagaatccgttaagaagagaag 2866
 |||||||
 DB 121 CACTTGAGTATATGCAATCTGTCCTTTGTGCAAAAAGATCCGTGAAGAGAGAGAG 180
 |||||||
 QY 2867 ctcatgctgaagcaatgttgggtgaaaaataataatgaagcgaggatctcgaagcagc 2926
 |||||||
 DB 181 CTCATGCTAGGCAATGTTTGGTGAATAATTAATGAAGCGGAGATCTGAAGCAGC 240
 |||||||
 QY 2927 atgcaactgttagtgaaaaattatgtctctcttaccagaagatgttggctcatataca 2986
 |||||||
 DB 241 ATGCAGCTGTTAGTGAATAATTTATGCTCTTTTACCAAGATGTTGTTCCATATACAA 300
 |||||||
 QY 2987 ttcaacttttggacatgaccagatatagttcaagaagatagatgaacaaactaaag 3046
 |||||||
 DB 301 TTCACTTTTGGCAATGACCCAGATATGCTAAAGTACAGAGATATGAACAACCTTAAG 360
 |||||||
 QY 3047 atgttaagaagaatgttgggttgccttgcgaataataatgaagcttaaaatgaataaca 3106
 |||||||
 DB 361 ATGTTAAAGAAATGCTTTGGTGTCTGGAATAATTAATGAGCTAAATAATGAATAACA 420
 |||||||
 QY 3107 gtcaagctttatcagaagaatgttgaagaataattaaacaacaagaagatgacaaagag 3166
 |||||||
 DB 421 GTCAAGCTTTTATCAGAAATGATGTAATAATTTAACAACAAGATGCCCAAGGAC 480
 |||||||
 QY 3167 cagatgatgcacaaatgaagaagaactgtacactgtgtgtgtatgttgcacatataca 3226
 |||||||
 DB 481 CAGATGATGCACAAATGAATGAATAAATCTGTACACTGTGTGATGTGCCATGATATCA 540
 |||||||
 QY 3227 tcatgtcaagaagtaactacatcacagtttggaaatcctctaagaagccggttaccaca 3282
 |||||||
 DB 541 TCATGTCAAAAGAGTACTACATACAGTTTGAATCTCTTAAAGACCCGGTACTACCA 596
 |||||||

RESULT 6
 BG387646 802 bp mRNA EST 12-MAR-2001
 LOCUS 602412496f1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4521095 5',
 DEFINITION mRNA sequence.
 ACCESSION BG387646
 VERSION BG387646.1 GI:13281092
 KEYWORDS EST.
 SOURCE human.

Matches	700; Conservative	0; Mismatches	21; Indels	4; Gaps	2;
QY 3352	tcaaaagtactacatcacagtttggatctccctaaagaccggtactaccagcttc	3291			
Db 1	TCAAAGAGTACTACTACAG-TTGGAAATCTCTTAAGAACCGGACTACAGCTGGTTTC	59			
QY 3292	ttcactcaacctggaagaatttagtgaacccaaaattatctgcctccctgaaatgaa	3351			
Db 60	TTCACTCAACCTGCACAAATTTTCAGTACACCAAAAATTTATCTGCTCTTGAATGAAA	119			
QY 3352	tcaattttcactccttggaaaaccttaaaaccaaactgtcttagaagctgttaacaagca	3411			
Db 120	TCATTTTTCACCTCTCGGAAAAACCTTAAACAACCAATGTTTTCAGAGCTGTAAACAGCA	179			
QY 3412	ctttcatcagcaggaagaacatcttaagccaaatcatcatcaggaatggaaacgttaagcaat	3471			
Db 180	CTTTATATAGCAGGCAACCAATCTTAGACCAAAATATATACAAATGGAACGTGAAGCAT	239			
QY 3472	gcaagcagcagctcacaatccaaagctcctcttggaaatlaaaaggggagcttgatagttct	3531			
Db 240	GCAAGCAGCAGCTCAAAATCCAAGCTCTCTGGAAGAAATAAAGGGAGGCTTGATAGTTCT	299			
QY 3532	gaaatggatcacagtggaaaatggaattacaacatgtcttcaaccttgcggggaaaaaa	3591			
Db 300	GAAATGGATCTACAGTGAAGAAATGCAATTTACAAATGTCACCTTTCACGCGGAAAAAA	359			
QY 3592	agtgcaagagagagacgactctgactctgttgaagtcgtgaatttggaaagccttagaagcag	3651			
Db 360	AGTGCAGAGAGAGACGACTCTGATCTTTAAAGTCTGTAATTTGGAGAACCTTAGAGCGAG	419			
QY 3652	aaaaaaagcccgctcacagaacagagagagaatagtgatgatalgtatgactgaagtgtg	3711			
Db 420	AAAAAAGCCCGCTCACAGAACAGAGAGAGAAATAGATGATGACTTACTCAAGTTTG	479			
QY 3712	gtacaggaagcagaaaccttaaaagcgatcaggaagtcgaaaaaagagccatacagcttca	3771			
Db 480	GTACAGGAACAGAAACCTTAAGGCGATGACGCGAAGCTCGGAAAGAGGCCATACGGCTTCA	539			
QY 3772	gaatctgtatgaacagcagcttgcccttgaggaaaaagagctcaagaagaatataatagaana	3830			
Db 540	GAACTGTGATGAACACACATGCGCTCAGGAAAAAGAGCGCTCAAAAGAGATATATTGAAACAA	599			
QY 3831	--tgaagtgacaggaatctgacgcgcgcaaaaaaaggtaaaagagcgccacccaaaact	3888			
Db 600	CTGAACGATGAACCAATATGTCGCCCAAAACAGGGGTACAAAGAGCGCGACCACAAAACCT	659			
QY 3889	cttggctgaggttacaccaaaagaagagcccaacatgaaaacctcttaaaaaaggaagcaaa	3948			
Db 660	CTTGCTGAGGATGACCAAAAAGAGAGCCACAGATGAATACTTCTATATAAGGAAGCAAA	719			
QY 3949	aaaaa 3953				
Db 720	CAAAA 724				
RESULT 3					
LOCUS BE873840	848 bp mRNA EST 20-OCT-2000				
DEFINITION	60148392p1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3886511 5',				
ACCESSION	BE873840				
VERSION	BE873840.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	NIH-MGC http://mgc.ncl.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email:cgapbs-remail.nih.gov				

Result No.	Score	Query Match	Length	DB	ID	Description
1	716	17.2	791	149	BF509252	UI-H-B14- BF509252 UI-H-B14-
2	663.4	15.9	734	175	BF5256731	602370939
3	621.2	14.9	848	141	BF873840	601483992
4	612.4	14.7	729	153	BF4332960	602495783
5	596	14.3	536	136	BF504550	h58405.x
6	591.4	14.2	802	153	BF387646	60242496
7	559.2	13.4	578	120	AM770571	h186f11.x
8	532.8	12.8	944	145	BF142489	601788573
9	503.8	12.1	608	32	AV685306	AV685306
10	495	11.9	683	136	BF532986	601235013
11	482.6	11.6	535	137	BF551003	767b03.x
12	478.4	11.5	480	23	AF1655429	ts88f06.x
13	474.4	11.4	608	149	BF471938	UI-M-BH3-
14	474.8	11.4	490	149	BF511856	UI-H-BH3-
15	468	11.2	888	155	BF571128	602591748
16	466.8	11.2	470	4	AA282190	z89b08.r
17	423	10.1	423	13	AF1680124	tw65b01.x
18	406.8	9.7	648	237	BF570606	601329907
19	392	9.4	434	116	AA483768	56376 MAR
20	388.4	9.3	391	10	AA630616	ac11e07.s
21	379.2	9.1	735	154	BF532020	602561049
22	377	9.0	445	142	BF548523	UI-M-BH3-
23	376.4	9.0	738	105	AL045878	DRF2P434J
24	368.2	8.8	595	143	BF043498	BF250017A
25	365	8.7	703	112	AM175091	f132b08.y
26	360.2	8.6	437	8	AA511220	vh5b10.r
27	360	8.6	486	13	AA393593	vz51f06.r
28	353	8.5	560	149	BF451878	u820g04.y
29	352.4	8.4	885	153	BF391029	602417722
30	351.8	8.4	429	145	BF189753	235446 MA
31	331.8	8.0	896	175	BF258248	602379670
32	329	7.9	408	11	AA755902	v004g01.r
33	326.8	7.8	701	154	BF473117	602515089
34	326.4	7.8	903	141	BE847511	uw27f05.y
35	321.6	7.7	902	169	BF789050	602104907
36	320.2	7.7	421	23	AF164566	v004g01.x
37	320.2	7.7	504	16	AF1120127	uh81h05.r
38	315.2	7.6	560	139	BF756042	210068 MA
39	311.4	7.5	728	32	AV710229	AV710229
40	310.6	7.4	662	172	BF016343	de59b11.y
41	309.4	7.4	883	155	BF5071195	602591130
42	302.6	7.3	751	251	AF2902024	RPECI-24-1
43	299.6	7.2	570	102	AF1794456	fc44f03.y
44	298.6	7.2	777	175	BF258462	602379941
45	297.6	7.1	560	23	AF166974	fc24a12.y

ALIGNMENTS

RESULT 1

BE509252 791 bp mRNA

LOCUS UI-H-B14-aow-c-07-0-UI-H-NCI CGAP_Sub8 Homo sapiens CDNA clone

DEFINITION IMAGE:3086220 3', mRNA sequence.

ACCESSION BF509252.1 GI:11592550

VERSION BF509252

KEYWORDS EST.

ORGANISM human.

SCIENCE Homo sapiens

Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 791)

NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE Tumor Gene Index

UNPUBLISHED (1997)

SUMMARIES

FEATURES

source

location/Qualifiers

1. 791

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NCI-CGAP_Sub8"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT7G


```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Title:	US-09-512-581-3
Perfect score:	4173
Sequence:	I atgctcatctcaagactag.....cacaaccaaaaaaaaaatgtg 4173

Scoring table:	IDENTITY_NUC
Score 100	Count 100

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

1:	Est1.*
2:	gb_est1.*
3:	gb_est2.*
4:	gb_est3.*
5:	gb_est4.*
6:	gb_est5.*
7:	gb_est6.*
8:	gb_est7.*
9:	gb_est8.*
10:	gb_est9.*
11:	gb_est10.*
12:	gb_est11.*
13:	gb_est12.*
14:	gb_est13.*
15:	gb_est14.*
16:	gb_est15.*
17:	gb_est16.*
18:	gb_est17.*
19:	gb_est18.*
20:	gb_est19.*
21:	gb_est20.*
22:	gb_est21.*
23:	gb_est22.*
24:	gb_est23.*
25:	gb_est24.*
26:	gb_est25.*
27:	gb_est26.*
28:	gb_est27.*
29:	gb_est28.*
30:	gb_est29.*
31:	gb_est30.*
32:	gb_est31.*
33:	gb_est32.*
34:	gb_est33.*
35:	gb_est34.*
36:	gb_est35.*
37:	gb_est36.*
38:	gb_est37.*
39:	gb_est38.*
40:	gb_est39.*
41:	gb_est40.*
42:	gb_est41.*
43:	gb_est42.*

14: em_esthnm10.*
 45: em_esthnm11.*
 46: em_esthnm12.*
 47: em_esthnm13.*
 48: em_esthnm14.*
 49: em_esthnm15.*
 50: em_esthnm16.*
 51: em_esthnm17.*
 52: em_esthnm18.*
 53: em_esthnm19.*
 54: em_esthnm20.*
 55: em_esthnm21.*
 56: em_esthnm22.*
 57: em_esthnm23.*
 58: em_esthnm24.*
 59: em_esthnm25.*
 60: em_esthnm26.*
 61: em_esthnm27.*
 62: em_esthnm28.*
 63: em_estin1.*
 64: em_estin2.*
 65: em_estin3.*
 66: em_estin4.*
 67: em_estin5.*
 68: em_estin6.*
 69: em_estin7.*
 70: em_estin8.*
 71: em_estin9.*
 72: em_estin10.*
 73: em_estin11.*
 74: em_estin12.*
 75: em_estin13.*
 76: em_estin14.*
 77: em_estin15.*
 78: em_estin16.*
 79: em_estin17.*
 80: em_estin18.*
 81: em_estin19.*
 82: em_estin20.*
 83: em_estin21.*
 84: em_estin22.*
 85: em_estin23.*
 86: em_estin24.*
 87: em_estin25.*
 88: em_estin26.*
 89: em_estin27.*
 90: em_estin28.*
 91: em_estin29.*
 92: em_estin30.*
 93: em_estin31.*
 94: em_estin32.*
 95: em_estin33.*
 96: em_estin34.*
 97: em_estin35.*
 98: em_estin36.*
 99: em_estin37.*
 100: em_estin38.*
 101: em_estin39.*
 102: em_estin40.*
 103: em_estin41.*
 104: em_estin42.*
 105: em_estin43.*
 106: em_estin44.*
 107: em_estin45.*
 108: em_estin46.*
 109: em_estin47.*
 110: em_estin48.*
 111: em_estin49.*
 112: em_estin50.*
 113: em_estin51.*
 114: em_estin52.*
 115: em_estin53.*
 116: em_estin54.*

QY 1324 PEEEEEE 1331
|||||

DB 551 PEEEEEE 558

RESULT 50

SQ7720

lamin B - mouse

C;Species: Mus musculus (house mouse)

C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 13-Aug-1999

C;Accession: S07720; S12488

R;Hoeger, T.H.; Krohne, G.; Franke, W.W.

Eur. J. Cell Biol. 47, 283-290, 1988

A;Title: Amino acid sequence and molecular characterization of murine lamin B as deduced

A;Reference number: S07720; MUID:89210899

A;Accession: S07720

A;Molecule type: mRNA

A;Residues: 1-587 <HOE>

A;Cross-references: EMBL:X16705

R;Hoeger, T.H.

submitted to the EMBL Data Library, October 1989

A;Reference number: S12488

A;Accession: S12488

A;Molecule type: mRNA

A;Residues: 1-530, 'I', 532-544, 'T', 546-552, 'E', 553-587 <HOE2>

A;Cross-references: EMBL:X16705; NID:g52868; PIDN:CAA34677.1; PID:g52869

C;Superfamily: cytoskeletal keratin

C;Keywords: nucleus

Query Match

0.6%; Score 8; DB 2; Length 587;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
|||||

DB 552 PEEEEEE 559

Search completed: September 25, 2001, 10:06:54

Job time: 160 sec

A:Molecule type: DNA
A:Residues: 1-568 <STO>
A:Cross-references: GB:BA000008; NID:g8578870; PIDN:BAA98706.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
C:Superfamily: proline--trRNA ligase

Query Match 0.6%; Score 8; DB 2; Length 568;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 188 GDTVSQEL 195
|||||
DB 483 GDTVSQEL 490

RESULT 47
F72070
C:Proline--trRNA ligase (EC 6.1.1.15) - Chlamydomophila pneumoniae (strains CWL029 and AR39)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: F72070; C81598
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: F72070
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <ARN>
A:Cross-references: GB:AE001363; NID:g4376785; PIDN:RAD18640.1; PID:g437678
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: C81598
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <REA>
A:Cross-references: GB:AE002185; GB:AE002161; NID:g7189170; PIDN:AAF38117.1; PID:g718918
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: proS; CP0254
C:Superfamily: proline--trRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 0.6%; Score 8; DB 2; Length 568;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 188 GDTVSQEL 195
|||||
DB 483 GDTVSQEL 490

RESULT 48
A96766
unknown protein F2p28 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96766
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Andersen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: A96766
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-575 <STO>
A:Cross-references: GB:AE005173; NID:g7109486; PIDN:AAF36750.1; GSPDB:GN00141
C:Genetics:
A:Gene: F2P9.28
A:Map position: 1

Query Match 0.6%; Score 8; DB 2; Length 575;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||
DB 188 EEEEEER 195

RESULT 49
VEHULB
Lamin B1 - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1991 #sequence_revision 15-Nov-1996 #text_change 10-Dec-1999
C:Accession: A34707; A57124
R:Pollard, K.M.; Chan, E.K.L.; Grant, B.J.; Sullivan, K.F.; Tan, E.M.; Glass, C.A. Mol. Cell. Biol. 10, 2164-2175, 1990
A:Title: In vitro posttranslational modification of lamin B cloned from a human T-cell
A:Reference number: A34707; MUID:90220602
A:Accession: A34707
A:Molecule type: mRNA
A:Residues: 1-586 <POL>
A:Cross-references: GB:M34458; NID:g186877; PIDN:AAA36162.1; PID:g307106
R:Lin, F.; Worman, H.J. Genomics 27, 230-236, 1995
A:Title: Structural organization of the human gene (LMNB1) encoding nuclear lamin B1.
A:Reference number: A57124; MUID:96044426
A:Accession: A57124
A>Status: translation not shown; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-586 <RES>
A:Cross-references: GB:L37747; NID:g576838; PIDN:NAC37575.1; PID:g576840
C:Comment: The lamins (A, B, and C) contains several alpha-helical domains capable of rates with lamin dissociation; it does not reform until telophase, when the lamins ar
C:Genetics:
A:Gene: GDB:LMNB1
A:Cross-references: GDB:512284; OMIM:150340
A:Map position: 5q23.3-5q31.1
A:Introns: 120/2; 172/3; 214/3; 271/3; 313/3; 387/2; 462/3; 497/3; 537/3; 573/3
C:Function:
A:Description: structural component of the nuclear lamina, a fibrous meshwork on the
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; lipoprotein; membrane protein; methylated carboxyl end; nucl
F:2-583/Product: lamin B1 #status predicted <MAT>
F:2-34/Domain: head <HED>
F:35-390/Domain: rod <ROD>
F:35-71/Region: coil 1A
F:82-240/Region: coil 1B
F:244-390/Region: coil 2
F:391-583/Domain: tail <END>
F:415-418/Region: nuclear location signal
F:583/Binding site: farnesyl (Cys) (covalent) #status predicted
F:583/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predict

Query Match 0.6%; Score 8; DB 1; Length 586;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 42

T04653
receiver-like protein 4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 03-Nov-2000
C:Accession: T04653; T52035
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15263
A:Accession: T04653
A:Molecule type: DNA
A:Residues: 1-552 <BEV>
A:Cross-references: EMBL:AL021636
A:Experimental source: cultivar Columbia; BAC clone F10N7
R:Buchholz, G.; Keitel, C.; Kircher, S.; Kudla, J.; Lohrmann, J.; Sweere, U.; Schaefer,
submitted to the EMBL Data Library, May 1998
A:Description: Nuclear-localized receiver-like proteins are differentially expressed in
A:Reference number: Z25912
A:Accession: T52035
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-552 <BU>
A:Cross-references: EMBL:AJ005195; PIDN:CAA06432.1
C:Genetics:
A:Gene: ARP4
A:Map position: 4
A:Introns: 43/1; 94/1; 207/1; 232/3; 517/3
A:Note: F10N7.270
C:Superfamily: response regulator homology
F:19-129/domain: response regulator homology <RRH>

Query Match 0.6%; Score 8; DB 2; Length 552;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||
DB 164 EEEEEER 171

RESULT 43

T49788
related to merozoite surface antigen 2 [imported] - Neurospora crassa
N:Alternate names: protein B9J10.250
C:Species: Neurospora crassa
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49788
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49788
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-562 <SCH>
A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.250
A:Experimental source: BAC clone B9J10; strain OR74A
C:Genetics:
A:Gene: NCSP:B9J10.250
A:Map position: 6
A:Introns: 386/2; 477/2; 527/1

Query Match 0.6%; Score 8; DB 2; Length 562;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||
DB 100 EEEEEER 107

RESULT 44

S37241
legumin B - fava bean
C:Species: Vicia faba (fava bean)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C:Accession: S46503; S37241
R:Heim, U.; Baumeister, H.; Mobus, U.
Plant Mol. Biol. 25, 131-135, 1994
A:Title: The legumin gene family: a reconstructed Vicia faba legumin gene encoding a
A:Reference number: S46503; MUID:94272010
A:Accession: S46503
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-564 <HE2>
A:Cross-references: EMBL:Z26489; NID:g403335; PIDN:CAA81262.1; PID:g403336
C:Genetics:
A:Introns: 178/3; 439/3
C:Superfamily: glycinin

Query Match 0.6%; Score 8; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||
DB 286 EEEEEER 293

RESULT 45

A72329
general secretion pathway protein E - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72329
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.N.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: A72329
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <ARN>
A:Cross-references: GB:AE001750; GB:AE000512; NID:g4981356; PIDN:AAD35919.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0837
C:Superfamily: secretion protein xcpR

Query Match 0.6%; Score 8; DB 2; Length 566;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||
DB 167 EEEEEER 174

RESULT 46

H86552
prolyl tRNA synthetase [imported] - Chlamydomophila pneumoniae (strain J138)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: H86552
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: H86552
A>Status: preliminary

A:Reference number: Z14734; MUID:97305956

A:Accession: T10830

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-513 <PRE>

A:Cross-references: EMBL:AE000105; NID:g2182706; PID:g2182709

C:Genetics:

A:Gene: nlfk2

A:Genome: plasmid pNGR234a

C:Superfamily: dinitrogenase beta chain

C:Keywords: ATP; iron-sulfur protein; metalloprotein; molybdenum; nitrogen fixation; ox

Query Match 0.6%; Score 8; DB 2; Length 513;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 LRLTTTIL 858

|||||

DB 481 LRLTTTIL 488

RESULT 38

T00828

C:Species: Arabidopsis thaliana

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Mar-1999

C:Accession: T00828

R:de la Bastide, M.; Hameed, A.; Gnoj, L.; Jensen, K.; Shohdy, N.; Gottesman, T.; Habert

McCombie, W.R.

submitted to the EMBL Data Library, January 1999

A:Description: A. thaliana BAC T13L16 from chromosome IV, top arm.

A:Reference number: Z14205

A:Accession: T00828

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-513

A:Cross-references: EMBL:AC003952; NID:g2708736; PID:g2708738

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Introns: 51/1; 57/1; 255/1; 283/3; 387/2; 411/2

A>Note: T13L16.2

Query Match 0.6%; Score 8; DB 2; Length 513;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332

|||||

DB 159 EEEEEER 166

RESULT 39

A56110

C:Species: Homo sapiens

C:Date: 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 05-Nov-1999

C:Accession: A56110

R:Jackman, J.K.; Motto, D.G.; Sun, Q.; Tanemoto, M.; Turck, C.W.; Peltz, G.A.; Koretzky,

J. Biol. Chem. 270, 7029-7032, 1995

A:Title: Molecular cloning of SLP-76, a 76-kDa tyrosine phosphoprotein associated with c

A:Reference number: A56110; MUID:95221345

A:Accession: A56110

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-533 <JAC>

A:Cross-references: GB:U20158; NID:g806765; PID:AA50135.1; PID:g806766

C:Genetics:

A:Gene: GDB:LCP2; SLP-76

A:Cross-references: GDB:1230199; OMIM:601603

A:Map position: 5q33.1-5qter

C:Superfamily: SH2 homology
C:Keywords: phosphoprotein
F:422-520/Domain: SH2 homology <SH2>

Query Match 0.6%; Score 8; DB 2; Length 533;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 SSPLPGKK 1197

|||||

DB 297 SSPLPGKK 304

RESULT 40

I36911

C:Species: Aotus trivirgatus

C:Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Jun-1999

C:Accession: I36911

R:Tseng, H.; Green, H.

Mol. Biol. Evol. 6, 460-468, 1989

A:Title: The involucrin gene of the owl monkey: origin of the early region.

A:Reference number: I36911; MUID:90014142

A:Accession: I36911

A:Molecule type: DNA

A:Residues: 1-544 <RES>

A:Cross-references: GB:M25313; NID:gl76557; PIDN:AAA35375.1; PID:gl76558

C:Comment: During the terminal differentiation of keratinocytes, this protein from th

linked envelope under the plasma membrane.

C:Superfamily: involucrin

C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

F:153-501/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KE]-[LH]-[PL]-E-Q)

Query Match 0.6%; Score 8; DB 1; Length 544;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 SQELDTV 199

|||||

DB 16 SQELDTV 23

RESULT 41

B44841

C:Species: Xenopus laevis

C:Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 12-Apr-1995

C:Accession: B44841

R:Charnas, L.R.; Szaro, B.G.; Gainer, H.

J. Neurosci. 12, 3010-3024, 1992

A:Title: Identification and developmental

A:Reference number: A44841; MUID:92356194

A:Accession: B44841

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-544 <CHA>

A:Experimental source: brain

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIN:110225, NCBIP:110226)

C:Superfamily: cytoskeletal keratin

Query Match 0.6%; Score 8; DB 2; Length 544;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1330

|||||

DB 462 APEEEEE 469

Hypertension 14, 435-444, 1989
A>Title: Molecular cloning of chromogranin A from rat pheochromocytoma cells.
A:Reference number: A60746; MUID:90007662
A:Accession: A60746
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-462 <P>
A:Cross-references: GB:AF145445; NID:g5163367; PIDN:AAD40652.1; PID:g5163368
R:Iacangelo, A.; Okayama, H.; Eiden, L.E.
FBBS Lett. 227, 115-121, 1988
A>Title: Primary structure of rat chromogranin A and distribution of its mRNA.
A:Reference number: S00291; MUID:88112232
A:Accession: S00291
A:Molecule type: mRNA
A:Residues: 1-107, 'OQOQ', 108-462 <IAC>
A:Cross-references: EMBL:X06832; NID:g55950; PIDN:CAA29988.1; PID:g55951
R:Hutton, J.C.; Nielsen, E.; Kastern, W.
FBBS Lett. 236, 269-274, 1988
A>Title: The molecular cloning of the chromogranin A-like precursor of beta-granin and F
A:Reference number: S02543; MUID:88312980
A:Accession: S02543
A:Molecule type: mRNA
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Residues: 19-107, 'OQOQ', 108-462 <HUT>
A>Note: part of this sequence, including the amino end of beta-granin, was determined by
C:Superfamily: chromogranin A
C:Keywords: glycoprotein; phosphoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-462/Product: chromogranin A #status predicted <MAT>
F:92-107/Region: glutamine-rich
F:417-419/Region: cell attachment (R-G-D) motif
F:355-56/Disulfide bonds: #status predicted
F:68,147,216,349,358,386,408,409,413/Binding site: phosphate (Ser) (covalent) #status pr
F:189/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.6%; Score 8; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
DB 340 EEEEEER 347
|||||||

RESULT 34
A39868
chromogranin A precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A39868
R:Wu, H.J.; Rozansky, D.J.; Farmer, R.J.; Gill, B.M.; O'Connor, D.T.
J. Biol. Chem. 266, 13130-13134, 1991
A>Title: Structure and function of the chromogranin A gene. Clues to evolution and tissu
A:Reference number: A39868; MUID:91302337
A:Accession: A39868
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-463 <WUA>
A:Cross-references: GB:M64278; NID:g192723; PIDN:AAA37457.1; PID:g192724
C:Superfamily: chromogranin A
F:92-116/Region: glutamine-rich

Query Match 0.6%; Score 8; DB 1; Length 463;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
DB 341 EEEEEER 348
|||||||

RESULT 35
G71231
probable replication factor C subunit - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: G71231
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermo
A:Reference number: A71000; MUID:98344137
A:Accession: G71231
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-468 <KAW>
A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29182.1; PID:g3256499
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0113
C:Superfamily: replication factor C large chain

Query Match 0.6%; Score 8; DB 2; Length 468;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
DB 436 EEEEEER 443
|||||||

RESULT 36
F84558
hypothetical protein At2g17970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84558
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Taiton,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: A84420; MUID:20083487
A:Accession: F84558
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-507 <STO>
A:Cross-references: GB:AE002093; NID:g4406821; PIDN:AAD20129.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g17970
A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 507;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
DB 153 EEEEEER 160
|||||||

RESULT 37
T10830
nitrogenase (EC 1.18.6.1) molybdenum-iron protein beta chain - Rhizobium sp. (strain
C:Species: Rhizobium sp.
A:Variety: strain NGR234
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 19-Jan-2001
C:Accession: T10830
R:Freiberg, C.; Fellay, R.; Bairoch, A.; Broughton, W.J.; Rosenthal, A.; Perret, X.
Nature 387, 394-401, 1997
A>Title: Molecular basis of symbiosis between Rhizobium and legumes.

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625
 A:Accession: H86683
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-437 <STO>
 A:Cross-references: GB:AE005176; NID:g12723351; PIDN:AAK04570.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: pil37

Query Match 0.6%; Score 8; DB 2; Length 437;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1200 KRDSDLV 1207
 |||||
 DB 73 KRDSDLV 80

RESULT 30

E70046 iron transport system homolog yvrA - Bacillus subtilis

C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: E70046
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y.M.; Odawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E70046
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-442 <KUN>
 A:Cross-references: GB:Z99120; GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15321.1; PI
 A:Experimental source: strain 168
 C:Genetics:

A:Gene: yvrA
 C:Superfamily: Bacillus subtilis probable iron transport system yvrA; ATP-binding cassette
 C:Keywords: ATP; nucleotide binding; P-loop
 F:16-214/Domain: ATP-binding cassette homology <ABC>
 F:33-40/Region: nucleotide-binding motif A (P-loop)

Query Match 0.6%; Score 8; DB 1; Length 442;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 KDLDLIK 514
 |||||
 DB 174 KDLDLIK 181

RESULT 31

A37386

dnab protein homolog - Chlamydia trachomatis (serotype D) plasmid pCHL1

C:Species: Chlamydia trachomatis
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
 C:Accession: A37386
 R:Comanducci, M.; Ricci, S.; Cevenini, R.; Ratti, G.

Plasmid 23, 149-154, 1990
 A:Title: Diversity of the Chlamydia trachomatis common plasmid in biovars with differ
 A:Reference number: A37386; MUID:90301796
 A:Accession: A37386
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-451 <COM>
 A:Cross-references: GB:J03321; NID:g144462; PIDN:AAA91569.1; PID:g144465
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: phage P22 gene 12 protein

Query Match 0.6%; Score 8; DB 2; Length 451;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1169 RIKGRDLS 1176
 |||||
 DB 145 RIKGRDLS 152

RESULT 32

S01921 dnaB protein homolog - Chlamydia trachomatis plasmids

C:Species: Chlamydia trachomatis
 C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 15-Oct-1999
 C:Accession: S01921; S01023; S01022
 R:Comanducci, M.; Ricci, S.; Ratti, G.
 Mol. Microbiol. 2, 531-538, 1988
 A:Title: The structure of a plasmid of Chlamydia trachomatis believed to be required
 A:Reference number: S01920; MUID:89013895

A:Accession: S01921
 A:Molecule type: DNA
 A:Residues: 1-451 <COM>
 A:Cross-references: EMBL:X07547; NID:g40730; PIDN:CAA30421.1; PID:g40732
 A:Experimental source: plasmid pCHL1
 R:Hatt, C.; Ward, M.E.; Clarke, I.N.
 Nucleic Acids Res. 16, 4053-4067, 1988
 A:Title: Analysis of the entire nucleotide sequence of the cryptic plasmid of Chlamydia
 A:Reference number: S00788; MUID:88233998

A:Accession: S01023
 A:Molecule type: DNA
 A:Residues: 88-451 <HAT>
 A:Cross-references: EMBL:X06707
 A:Experimental source: plasmid pLGV440
 A:Accession: S01022
 A:Molecule type: DNA
 A:Residues: 1-78, 'IWTNRS', 86 <HA2>
 A:Cross-references: EMBL:X06707
 A:Experimental source: plasmid pLGV440
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: phage P22 gene 12 protein

Query Match 0.6%; Score 8; DB 2; Length 451;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1169 RIKGRDLS 1176
 |||||
 DB 145 RIKGRDLS 152

RESULT 33

A60746

chromogranin A precursor - rat

N:Contains: beta-granin; pancreastatin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: A60746; S00291; S02543
 R:Parmer, R.J.; Koop, A.H.; Handa, M.T.; O'Connor, D.T.

A:Reference number: A69250; MUID:98049343
A:Accession: B69436
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <KLE>
A:Cross-references: GB:AF000999; GB:AF000782; NID:g2689322; PIDN:AAB89749.1; PID:g264906
C:Superfamily: rat acidic ribosomal protein P0

Query Match 0.6%; Score 8; DB 2; Length 339;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db 317 PEEEEEE 324

RESULT 25
G72775
hypothetical protein APE0193 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72775
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339
A:Accession: G72775
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-341 <KAW>
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79105.1; PID:d1042881; PID:g510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0193

Query Match 0.6%; Score 8; DB 2; Length 341;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 981 EKLKSLLP 988
Db 325 EKLKSLLP 332

RESULT 26
B71604
rRNA methylase (SpoU family) (OO, TP) PFB0855c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: B71604
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Perlea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: B71604
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-351 <GAR>
A:Cross-references: GB:AF001421; GB:AF001362; NID:g3845293; PIDN:AAC71960.1; PID:g384529
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0855c

Query Match 0.6%; Score 8; DB 2; Length 351;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1270 LKEDILEN 1277
Db 305 LKEDILEN 312

RESULT 27
S06324
dnaB protein homolog - Chlamydia trachomatis plasmid pCtT1
C:Species: Chlamydia trachomatis
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 11-Jan-2000
C:Accession: S06324
R:Striprakash, K.S.; MacAvoy, E.S. Nucleic Acids Res. 15, 10596, 1987
A:Title: A gene for dnaB like protein in chlamydial plasmid.
A:Reference number: S06324; MUID:88096599
A:Accession: S06324
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-398 <SRI>
C:Genetics:
A:Genome: plasmid
C:Superfamily: phage P22 gene 12 protein
C:Keywords: DNA binding

Query Match 0.6%; Score 8; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1169 RIKGRLDS 1176
Db 145 RIKGRLDS 152

RESULT 28
T15827
hypothetical protein C53C9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T15827
R:Bentley, D. submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid C53C9.
A:Reference number: Z18413
A:Accession: T15827
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-418 <BEN>
A:Cross-references: EMBL:U28734; NID:g861255; PID:g861256; PIDN:AAB52603.1; GSPDB:GNO
A:Experimental source: strain Bristol N2; clone C53C9
C:Genetics:
A:Gene: CESP:C53C9.2
A:Map position: X
A:Introns: 25/3; 187/1; 310/2; 380/3; 405/3

Query Match 0.6%; Score 8; DB 2; Length 418;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db 369 PEEEEEE 376

RESULT 29
H86683
prophage p11 protein 37, capsid protein [imported] - Lactococcus lactis subsp. lactis
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: H86683
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh Genome Res. in press, 2001

QY 1324 PEEEEEE 1331
|||||||
DB 226 PEEEEEE 233

RESULT 20
F84923
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84923
R:Jilin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84923
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <STO>
A:Cross-References: GB:AE002093; NID:g5729705; PIDN:AAD48512.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g48120
A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 313;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||||
DB 80 EEEEEER 87

RESULT 21
S58719
A:Title: Probable membrane protein YNL058c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N2433; hypothetical protein YNL1621
C:Species: Saccharomyces cerevisiae
C:Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 21-Jul-2000
C:Accession: S58719; S62986
R:Bergez, P.; Doignon, F.; Crouzet, M.
Yeast 11, 967-974, 1995
A:Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome XIV
A:Reference number: S58711; MUID:96021608
A:Accession: S58719
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-316 <BER>
A:Cross-References: EMBL:U12141; NID:gl1314216; PIDN:AAA99653.1; PID:g994828
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1994
R:Bergez, P.; Doignon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62975
A:Accession: S62986
A:Molecule type: DNA
A:Residues: 1-316 <BEW>
A:Cross-References: EMBL:Z71334; NID:gl1301928; PID:gl1301929; MIPS:YNL058C
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 14L
C:Keywords: transmembrane protein
F:95-111/Domain: transmembrane #status predicted <TMW>

Query Match 0.6%; Score 8; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
|||||||

DB 237 EEEEEER 244

RESULT 22
JC6146
C:Card protein - Myxococcus xanthus
C:Species: Myxococcus xanthus
C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 08-Oct-1999
C:Accession: JC6146
R:Nicolas, F.J.; Cayuela, M.L.; Martinez-Argudo, I.M.; Ruiz-Vazquez, R.M.; Murillo, F.
Proc. Natl. Acad. Sci. U.S.A. 93, 6881-6885, 1996
A:Title: High mobility group I(Y)-like DNA-binding domains on a bacterial transcript
A:Reference number: JC6146; MUID:96293442
A:Accession: JC6146
A:Molecule type: DNA
A:Residues: 1-316 <NIC>
A:Cross-References: EMBL:Z56280; NID:gl022325; PIDN:CAA91224.1; PID:gl022328
C:Comment: This protein is an eukaryotic transcription factor that is involved in bot
C:Genetics:
A:Gene: card
C:Keywords: leucine zipper; transcription factor

Query Match 0.6%; Score 8; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1289 KGRPPKP 1296
|||||||
DB 257 KGRPPKP 264

RESULT 23
T18283
A:Title: Hypothetical protein G5 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18283
R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.
Genetics 148, 1117-1125, 1998
A:Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Dd
A:Reference number: Z14684; MUID:98198836
A:Accession: T18283
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-325 <RIE>
A:Cross-References: EMBL:U00796; NID:g2702254; PID:g2702258; PIDN:AAC18634.1
C:Genetics:
A:Introns: 85/1

Query Match 0.6%; Score 8; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
|||||||
DB 159 PEEEEEE 166

RESULT 24
B69436
LSU ribosomal protein L10E (rpl10E) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 13-Aug-1999
C:Accession: B69436
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: S13625; MUID:91160746
A:Accession: S13625
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-233 <AOK>
A:Cross-references: GB:D90251; NID:93135096; PIDN:BAA14291.1; PID:g2220291
A:Accession: S18497
A:Molecule type: protein
A:Residues: 115-146,'X',148-157,'X',159-162 <AOK1>
R:Hashimoto, Y.; Nagaoaka, I.; Yamashita, T.
Biochim. Biophys. Acta 1203, 236-242, 1993
A:Title: Purification of the antibacterial fragments of guinea-pig major basic protein.
A:Reference number: S40491; MUID:94092714
A:Accession: S40491
A:Molecule type: protein
A:Residues: 115-233 <HAS>
C:Superfamily: eosinophil major basic protein precursor; C-type lectin homology
C:Keywords: antibiotic; chondroitin sulfate proteoglycan; cytotoxin; glycoprotein
F:11-16/Domain: signal sequence #status predicted <SIG>
F:17-114/Domain: propeptide #status predicted <PRO>
F:115-233/Product: eosinophil major basic protein 1 #status experimental <MAT>
F:116-231/Domain: C-type lectin homology <LCH>
F:24-25/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:67/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:134-231,208-223/Disulfide bonds: #status predicted

Query Match 0.6%; Score 8; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 1324 PEEEEEE 1331
DB 73 PEEEEEE 80
|||||||

RESULT 17
A:33729
C:Species: Oncorhynchus keta (chum salmon)
C:Date: 20-Sep-1991 #sequence_revision 20-Sep-1991 #text_change 20-Jun-2000
A:Accession: A23729; B49762
R:Takayama, Y.; Rand-Weaver, M.; Kawauchi, H.; Ono, M.
Mol. Endocrinol. 5, 778-786, 1991
A:Title: Gene structure of chum salmon somatolactin, a presumed pituitary hormone of the
A:Reference number: A23729; MUID:92017883
A:Accession: A23729
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-233 <TAK>
A:Cross-references: GB:D10638; NID:g222909; PIDN:BAA01485.1; PID:g222911
R:Takayama, Y.; Ono, M.; Rand-Weaver, M.; Kawauchi, H.
Gen. Comp. Endocrinol. 83, 366-374, 1991
A:Title: Greater conservation of somatolactin, a presumed pituitary hormone of the growth
A:Reference number: A49762; MUID:92038679
A:Accession: B49762
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-233 <TA2>
A:Cross-references: GB:D10640; NID:g222903; PIDN:BAA01487.1; PID:g222904
C:Superfamily: prolactin

Query Match 0.6%; Score 8; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QX 979 VSEKLLSL 986
DB 150 VSEKLLSL 157
|||||||

RESULT 18

RHHUT
thyroliberin precursor [validated] - human
N:Alternate names: prothyrotropin; thyrotropin-releasing hormone precursor; TRH
N:Contains: prothyroliberin; thyroliberin
C:Species: Homo sapiens (man)
C:Date: 22-Jun-1990 #sequence_revision 12-May-1995 #text_change 08-Dec-2000
A:Accession: A34550
R:Yamada, M.; Radovick, S.; Wondisford, F.E.; Nakayama, Y.; Weintraub, B.D.; Wilber, M.L. Endocrinol. 4, 551-556, 1990
A:Title: Cloning and structure of human genomic DNA and hypothalamic cDNA encoding hu
A:Reference number: A34550; MUID:91125361
A:Accession: A34550
A:Molecule type: DNA; mRNA
A:Residues: 1-242 <YAM>
A:Cross-references: GB:M63581; GB:M63582; NID:g190296; PIDN:AAA36480.1; PID:g190298;
C:Comment: The peptide hormone is released from the hypothalamus.
C:Genetics:
A:Gene: GDB:TRH
A:Cross-references: GDB:128072; OMIM:275120
A:Map position: 3pter-3qter
A:Introns: 71/1
A:Note: the first intron occurs before the initiator codon
C:Function:
A:Description: stimulates synthesis and secretion of thyrotropin from the anterior pi
C:Superfamily: thyroliberin precursor
C:Keywords: amidated carboxyl end; hormone; hypothalamus; polypeptide; pyroglutamic a
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-242/Product: prothyroliberin #status predicted <PTL>
F:84-86/Product: thyroliberin #status experimental <TL1>
F:114-116/Product: thyroliberin #status experimental <TL2>
F:135-137/Product: thyroliberin #status experimental <TL3>
F:152-154/Product: thyroliberin #status experimental <TL4>
F:186-188/Product: thyroliberin #status experimental <TL5>
F:227-229/Product: thyroliberin #status experimental <TL6>
F:84,114,135,152,186,227/Modified site: pyrrolidone carboxylic acid (Gln) (in mature
F:86,116,137,154,188,229/Modified site: amidated carboxyl end (Pro) (amide in mature

Query Match 0.6%; Score 8; DB 1; Length 242;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 169 EEEEEER 176
|||||||

RESULT 19
E84706
hypothetical protein At2g30280 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84706
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84706
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <STO>
A:Cross-references: GB:A8002093; NID:g2347194; PIDN:AAC16933.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g30280
A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A:Accession: B46315
A:Molecule type: DNA
A:Residues: 1-147 <DOO>
A:Cross-references: GB:M28744; NID:g333136; PIDN:AAA47015.1; PID:g333137
A:Superfamily: papillomavirus type 4 E4 protein
C:Keywords: early protein

Query Match 0.6%; Score 8; DB 1; Length 147;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
|||||||
DB 102 EEEEEER 109

RESULT 12
TQ5575
hypothetical protein F22K18.210 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
C:Accession: TQ5575
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15419
A:Accession: TQ5575
A:Molecule type: DNA
A:Residues: 1-204 <BEV>
A:Cross-references: EMBL:AL035356
A:Experimental source: cultivar Columbia; BAC clone F22K18
C:Genetics:
A:Map position: 4
A:Introns: 85/3; 116/1; 169/3
A:Note: F22K18.210
C:Superfamily: Arabidopsis hypothetical protein F22K18.210

Query Match 0.6%; Score 8; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
|||||||
DB 60 PEEEEEE 67

RESULT 13
JC1129
nonhistone chromosomal protein HMG-2 - chicken
N:Alternate names: high-mobility-group protein 2
C:Species: Gallus gallus (chicken)
C:Date: 05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change 23-Jul-1999
C:Accession: JC1129
R:Sparrow, D.B.; Wells, J.R.E.
Gene 114, 289-290, 1992
A:Title: Sequence of a cDNA encoding chicken high-mobility-group protein-2.
A:Reference number: JC1129; MUID:92290291
A:Accession: JC1129

A:Molecule type: mRNA
A:Residues: 1-207 <SPA>
A:Cross-references: GB:M80574; NID:g211928; PIDN:AAA48819.1; PID:g211929
C:Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology
C:Keywords: chromosomal protein; DNA binding; nucleus
F:6-83/Domain: HMG box homology <HMG1>
F:92-166/Domain: HMG box homology <HMG2>

Query Match 0.6%; Score 8; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331

DB 187 PEEEEEE 194
|||||||

RESULT 14
JC1114
high-mobility group protein 2 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 23-Jul-1999
C:Accession: JC1114
R:Davis, D.L.; Burch, J.B.E.
Gene 113, 251-256, 1992
A:Title: Isolation of a chicken HMG2 cDNA clone and evidence for an HMG2-specific 3'-
A:Reference number: JC1114; MUID:92241676
A:Accession: JC1114
A:Molecule type: mRNA
A:Residues: 1-207 <DAV>
A:Cross-references: GB:M83235; NID:g211926; PIDN:AAA48818.1; PID:g211927
C:Comment: The high mobility group proteins are among the most abundant nonhistone ch
C:Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology
C:Keywords: DNA binding; nucleus
F:6-83/Domain: HMG box homology <HMG1>
F:92-166/Domain: HMG box homology <HMG2>

Query Match 0.6%; Score 8; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1324 PEEEEEE 1331
|||||||
DB 187 PEEEEEE 194

RESULT 15
T49830
hypothetical protein B24H17.160 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49830
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49830
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <SCH>
A:Cross-references: EMBL:AL356815; GSPDB:GN00116; NCSP:B24H17.160
A:Experimental source: BAC clone B24H17; strain OR74A
C:Genetics:
A:Gene: NCSP:B24H17.160
A:Map position: 6
A:Introns: 38/1; 102/3; 128/3

Query Match 0.6%; Score 8; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1325 EEEEEER 1332
|||||||
DB 70 EEEEEER 77

RESULT 16
S13625
eosinophil major basic protein 1 precursor - guinea pig
C:Species: Cavia porcellus (guinea pig)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
C:Accession: S13625; S18497; S40491
R:Aoki, I.; Shindoh, Y.; Nishida, T.; Nakai, S.; Hong, Y.M.; Mio, M.; Saito, T.; Tasa
FEBS Lett. 279, 330-334, 1991
A:Title: Sequencing and cloning of the cDNA of guinea pig eosinophil major basic prot

Query Match 0.6%; Score 8; DB 2; Length 81;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 EEEEEER 1332
 |||||
 Db 19 EEEEEER 26

RESULT 8
 S27275
 GTP-binding regulatory protein gamma chain precursor - northern European squid
 N:Alternate names: G-protein gamma chain; transducin gamma chain
 C:Species: Loligo forbesi (northern European squid)
 C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 07-May-1999
 C:Accession: S27275; S17596; S25626
 R:Jott, J.S.; Ryba, N.J.P.; Pottinger, J.D.D.; Keen, J.N.; Carne, A.; Findlay, J.B.C.
 F:BS Lett. 312, 241-244, 1992
 A:Title: The gamma-subunit of the principal G-protein from squid (Loligo forbesi) photoreceptor
 A:Reference number: S27275; MUID:93050202
 A:Accession: S27275
 A:Molecule type: mRNA
 A:Residues: 1-87 <LOT>
 A:Cross-references: EMBL:215112; NID:g9509; PID:g9510
 A:Experimental source: retina
 R:Pottinger, J.D.D.; Ryba, N.J.P.; Keen, J.N.; Findlay, J.B.C.
 B:Chem. J. 279, 323-326, 1991
 A:Title: The identification and purification of the heterotrimeric GTP-binding protein from squid (Loligo forbesi)
 A:Reference number: S17596; MUID:92028830
 A:Accession: S17596
 A:Molecule type: protein
 A:Residues: 'LE', 29, 'T', 31-33, 'P', 35-41, 43-51, 'M', 53-59, 'D', 61-63, 'MK', 66, 'A', 68-70, 'X', 72-74, 'Y', 76-78, 'Q', 80-82, 'K', 84-86, 'R', 88-90, 'G', 92-94, 'S', 96-98, 'L', 100-102, 'I', 104-106, 'V', 108-110, 'F', 112-114, 'C', 116-118, 'H', 120-122, 'E', 124-126, 'D', 128-130, 'N', 132-134, 'K', 136-138, 'Q', 140-142, 'R', 144-146, 'G', 148-150, 'S', 152-154, 'L', 156-158, 'I', 160-162, 'V', 164-166, 'F', 168-170, 'C', 172-174, 'H', 176-178, 'E', 180-182, 'D', 184-186, 'N', 188-190, 'K', 192-194, 'Q', 196-198, 'R', 200-202, 'G', 204-206, 'S', 208-210, 'L', 212-214, 'I', 216-218, 'V', 220-222, 'F', 224-226, 'D', 228-230, 'N', 232-234, 'K', 236-238, 'Q', 240-242, 'R', 244-246, 'G', 248-250, 'S', 252-254, 'L', 256-258, 'I', 260-262, 'V', 264-266, 'F', 268-270, 'C', 272-274, 'H', 276-278, 'E', 280-282, 'D', 284-286, 'N', 288-290, 'K', 292-294, 'Q', 296-298, 'R', 300-302, 'G', 304-306, 'S', 308-310, 'L', 312-314, 'I', 316-318, 'V', 320-322, 'F', 324-326, 'D', 328-330, 'N', 332-334, 'K', 336-338, 'Q', 340-342, 'R', 344-346, 'G', 348-350, 'S', 352-354, 'L', 356-358, 'I', 360-362, 'V', 364-366, 'F', 368-370, 'C', 372-374, 'H', 376-378, 'E', 380-382, 'D', 384-386, 'N', 388-390, 'K', 392-394, 'Q', 396-398, 'R', 400-402, 'G', 404-406, 'S', 408-409
 F:85-87/Domain: heterotrimer; lipoprotein; membrane protein; prenylated cysteine; signal tra
 F:84/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 0.6%; Score 8; DB 2; Length 87;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 EEEEEER 1332
 |||||
 Db 12 EEEEEER 19

RESULT 9
 S00219
 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) 11k protein precursor - human
 N:Alternate names: complex III 11k protein; cytochrome bcl complex 11k protein; mitochondrion
 C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1988 #sequence_revision 22-Jul-1994 #text_change 21-Jul-2000
 C:Accession: S00219; I52694
 R:Ohta, S.; Goto, K.; Arai, H.; Kagawa, Y.
 F:BS Lett. 226, 171-175, 1987
 A:Title: An extremely acidic amino-terminal presequence of the precursor for the human m
 A:Reference number: S00219; MUID:88083627
 A:Accession: S00219
 A:Molecule type: mRNA
 A:Residues: 1-91 <OHT>
 A:Cross-references: EMBL:M36647; NID:g188564; PIDN:AAA36317.1; PID:g188565
 R:Jiliu, A.Y.; Bradner, R.C.
 C:Date: Res. 53, 2460-2465, 1993
 A:Title: Elevated expression of the human mitochondrial hinge protein gene in cancer.
 A:Reference number: I52694; MUID:93265436
 A:Accession: I52694
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-56 <RES>
 A:Cross-references: G:561826; NID:g385936; PIDN:AAD13930.1; PID:g4261630
 C:Genetics:

A:Gene: GDB:UQCRH
 A:Cross-references: GDB:141852
 A:Map position: 22q13-22q13
 C:Superfamily: ubiquinol--cytochrome-c reductase 11k protein; ubiquinol--cytochrome-c
 C:Keywords: mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory cha
 F:1-13/Domain: transit peptide (mitochondrion) #status predicted <TRP>
 F:14-91/Product: ubiquinol--cytochrome-c reductase 11k protein #status predicted <MAT
 F:23-91/Domain: ubiquinol--cytochrome-c reductase 11k protein homology <U11>
 F:37-81.53-67/Disulfide bonds: #status predicted

Query Match 0.6%; Score 8; DB 1; Length 91;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
 |||||
 Db 16 PEEEEEE 23

RESULT 10
 A55819
 nonhistone chromosomal protein CHMG-I - midge (Chironomus tentans)
 N:Alternate names: high mobility group protein I/Y homolog
 C:Species: Chironomus tentans
 C>Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 21-Jul-2000
 C:Accession: A55819; S47253; S47252
 R:Claus, P.; Schulze, E.; Wisniewski, J.R.
 J. Biol. Chem. 269, 33042-33048, 1994
 A:Title: Insect proteins homologous to mammalian high mobility group proteins I/Y (HM
 A:Reference number: A55819; MUID:95105193
 A:Accession: A55819
 A:Molecule type: mRNA
 A:Residues: 1-99 <CLA>
 A:Cross-references: EMBL:Z36898; NID:g534932; PIDN:CAA85365.1; PID:g534933
 A:Experimental source: clone pCM126
 R:Claus, P.; Schulze, E.; Wisniewski, J.R.
 submitted to the EMBL Data Library, August 1994
 A:Description: Insect proteins homologous to mammalian high mobility group proteins I
 A:Reference number: S47252
 A:Accession: S47253
 A:Molecule type: DNA
 A:Residues: 1-91, 'A', 93-99 <CL2>
 A:Cross-references: EMBL:Z36897; NID:g534886; PID:g534887
 A:Experimental source: clone pCWG106
 C:Genetics:
 A:Introns: 15/3; 76/1
 A:Note: suggested by Southern blot analysis in Ref A55819 to be single copy gene
 C:Keywords: chromosomal protein; DNA binding
 F:8-13/Region: DNA-binding motif (K/R-G-R-G-R-P)
 F:55-60/Region: DNA-binding motif (K/R-G-R-G-R-P)
 F:75-80/Region: DNA-binding motif (K/R-G-R-G-R-P)

Query Match 0.6%; Score 8; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1373 KGRGRPSK 1380
 |||||
 Db 55 KGRGRPSK 62

RESULT 11
 B46315
 E4 protein - human papillomavirus type 4
 C:Species: human papillomavirus type 4
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
 C:Accession: B46315
 R:Doorbar, J.; Coneron, I.; Gallimore, P.H.
 Virology 172, 51-62, 1989
 A:Title: Sequence divergence yet conserved physical characteristics among the E4 prot
 A:Reference number: A46315; MUID:89370332

RESULT 5

SQ6005
N:Myosin alpha heavy chain, cardiac muscle [similarity] - rat
N:Alternate names: alpha-myosin heavy chain
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
C:Accession: S06005; S07535; A20971; A02988; I53305
R:McNally, E.M.; Gianola, K.M.; Leinwand, L.A.
Nucleic Acids Res. 17, 7527-7528, 1989
A:Title: Complete nucleotide sequence of full length cDNA for rat alpha cardiac myosin H
A:Reference number: S06005; MUID:90016822
A:Accession: S06005
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-1938 <MC>
A:Cross-references: EMBL:X15938; NID:956654; PIDN:CAA34064.1; PID:g56655
R:McNally, E.M.; Kraft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
J. Mol. Biol. 210, 665-671, 1989
A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparison
A:Reference number: S07535; MUID:90133919
A:Accession: S07535
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1938 <MC2>
R:Mahdavi, V.; Chambers, A.P.; Nadal-Ginard, B.
Proc. Natl. Acad. Sci. U.S.A. 81, 2626-2630, 1984
A:Title: Cardiac alpha- and beta-myosin heavy chain genes are organized in tandem.
A:Reference number: A20971; MUID:84194059
A:Accession: A20971
A:Molecule type: protein
A:Residues: 1-12, 'AP', 14-45, 'A', 47-50, 'AP', 53-81, 'E', 83-86, 'Q', 88-109, 111-133, 'H', 135-16
R:Mahdavi, V.; Periasamy, M.; Nadal-Ginard, B.
Nature 297, 659-664, 1982
A:Title: Molecular characterization of two myosin heavy chain genes expressed in the adu
A:Reference number: A02988; MUID:82220036
A:Accession: A02988
A:Molecule type: mRNA
A:Residues: 1512-1574, 'S', 1576-1721, 'N', 1723-1851, 'N', 1853-1869, 'N', 1871-1933, 'I', 1935-1
A:Note: there are 10 or more myosin heavy chain genes in the rat, at least two of which
R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.
Eur. Heart J. 5, 181-191, 1984
A:Title: Cardiac myosin heavy chain isozymic transitions during development and under pa
A:Reference number: I53305; MUID:85179510
A:Accession: I53305
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1872-1933, 'I', 1935-1938 <RES>
A:Cross-references: GB:M32697; NID:9205596; PIDN:AAA41658.1; PID:g205597
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylate
F:87-767/Domain: myosin motor domain homology <MMOT>
F:177-184/Region: nucleotide-binding motif A (P-loop)
F:548-585/Region: actin binding #status predicted
F:656-678/Region: actin binding #status predicted
F:840-1938/Domain: coiled coil #status predicted <COL>
F:840-1280/Region: S2
F:1281-1938/Region: light meromyosin
F:128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:183/Binding site: ATP (Lys) #status predicted
F:696,706/Active site: Cys #status predicted

Query Match 0.6%; Score 9; DB 1; Length 1938;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKOLEV 565

Db 1528 EKIRKOLEV 1536
|||||

RESULT 6

I49464

alpha cardiac myosin heavy chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
C:Accession: I49464; I49463; I49462; I49461; I49604
R:Quinn-Laquer, B.K.; Kennedy, J.E.; Wei, S.J.; Beisel, K.W.
Genomics 13, 176-188, 1992
A:Title: Characterization of the allelic differences in the mouse cardiac alpha-myosi
A:Reference number: A38207; MUID:92250040
A:Accession: I49464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1938 <RES>
A:Cross-references: GB:M76601; NID:gl91623; PIDN:AAA37162.1; PID:gl91624
A:Accession: I49463
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-193, 'D', 195-837, 'S', 839-955, 'N', 957-1938 <RE2>
A:Cross-references: GB:M76600; NID:gl91621; PIDN:AAA37161.1; PID:gl91622
A:Accession: I49462
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1938 <RE3>
A:Cross-references: GB:M76599; NID:gl91619; PIDN:AAA37160.1; PID:gl91620
A:Accession: I49461
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-544, 'A', 546-1938 <RE4>
A:Cross-references: GB:M76598; NID:gl91617; PIDN:AAA37159.1; PID:gl91618
R:Gulick, J.; Subramaniam, A.; Neumann, J.; Robbins, J.
J. Biol. Chem. 266, 9180-9185, 1991
A:Title: Isolation and characterization of the mouse cardiac myosin heavy chain genes
A:Reference number: I49604; MUID:91225025
A:Accession: I49604
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-67 <RE5>
A:Cross-references: GB:M62404; NID:gl92609; PIDN:AAA37424.1; PID:gl92610
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop
F:98-768/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 0.6%; Score 9; DB 2; Length 1938;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 EKIRKOLEV 565

Db 1529 EKIRKOLEV 1537
|||||

RESULT 7

T47289
hypothetical protein T14K23.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47289
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M
Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24458
A:Accession: T47289
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-81 <NVA>
A:Cross-references: EMBL:AL132909
A:Experimental source: cultivar Columbia; BAC clone T14K23
C:Genetics:
A:Map position: 3
A:Introns: 26/2; 59/1
A:Note: T14K23.20

979 944 ALCADPKVRRRAHARQCL 962
980 468 ALCADPKVRRRAHARQCL 486
981 944 ALCADPKVRRRAHARQCL 962
982 468 ALCADPKVRRRAHARQCL 486
983 944 ALCADPKVRRRAHARQCL 962
984 468 ALCADPKVRRRAHARQCL 486
985 944 ALCADPKVRRRAHARQCL 962
986 468 ALCADPKVRRRAHARQCL 486
987 944 ALCADPKVRRRAHARQCL 962
988 468 ALCADPKVRRRAHARQCL 486
989 944 ALCADPKVRRRAHARQCL 962
990 468 ALCADPKVRRRAHARQCL 486
991 944 ALCADPKVRRRAHARQCL 962
992 468 ALCADPKVRRRAHARQCL 486
993 944 ALCADPKVRRRAHARQCL 962
994 468 ALCADPKVRRRAHARQCL 486
995 944 ALCADPKVRRRAHARQCL 962
996 468 ALCADPKVRRRAHARQCL 486
997 944 ALCADPKVRRRAHARQCL 962
998 468 ALCADPKVRRRAHARQCL 486
999 944 ALCADPKVRRRAHARQCL 962
1000 468 ALCADPKVRRRAHARQCL 486

ALIGNMENTS

RESULT 1
T00374
hypothetical protein KIAA0648 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00374
R:Fishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: 214142; MUID:98403880
A:Accession: T00374
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-851 <ISH>
A:Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PID:g3327110
A:Experimental source: brain
C:Genetics:
A:Note: KIAA0648

Query Match 1.4%; Score 19; DB 2; Length 851;
Best Local Similarity 100.0%; Pred. No. 3.9e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 944 ALCADPKVRRRAHARQCL 962
DB 468 ALCADPKVRRRAHARQCL 486
RESULT 2
A57573
telomeric repeat binding factor - human
C:Species: Homo sapiens (man)
C>Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 19-May-2000
C:Accession: A57573
R:Chong, L.; van Steensel, B.; Broccoli, D.; Erdjument-Bromage, H.; Hanish, J.; Tempst,
Science 270, 1663-1667, 1995
A>Title: A human telomeric protein.
A:Reference number: A57573; MUID:96099400
A:Accession: A57573
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-439 <CHO>
A:Cross-references: GB:U40705; NID:g2078442; PID:g1439542
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homolo
C:Keywords: DNA binding

F:375-428/Domain: myb DNA-binding repeat homology <MYB3>

Query Match 0.6%; Score 9; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEEEE 1331

DB 53 APEEEEEEE 61

RESULT 3

T45840

GTPase activating-like protein - Arabidopsis thaliana

N:Alternate names: protein F2K15.210

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T45840

R:Kieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Benck

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23015

A:Accession: T45840

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-554 <RIE>

A:Cross-references: EMBL:AL132956

A:Experimental source: cultivar Columbia; BAC clone F2K15

C:Genetics:

A:Map position: 3

A:Introns: 53/2; 98/3; 115/2; 133/2

A:Note: F2K15.210

Query Match 0.6%; Score 9; DB 2; Length 554;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 SSSSNPSP 1167

DB 11 SSSSNPSP 19

RESULT 4

T49562

neurofilament-H related protein [imported] - Neurospora crassa

N:Alternate names: protein B208.100

C:Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49562

R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49562

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-837 <SCH>

A:Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.100

A:Experimental source: BAC clone B208; strain OR74A

C:Genetics:

A:Gene: NCSP:B208.100

A:Map position: 6

Query Match 0.6%; Score 9; DB 2; Length 837;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1287 KGKRGPPK 1295

DB 695 KGKRGPPK 703

833	7	0.5	1545	2	T46645	sulfonyleurea recep	906	6	0.4	39	2	D83721	hypothetical prote
834	7	0.5	1546	2	T42728	sulfonyleurea recep	907	6	0.4	40	2	G82271	hypothetical prote
835	7	0.5	1558	2	T29253	hypothetical prote	908	6	0.4	41	2	T14350	tubulin beta-3 cha
836	7	0.5	1560	2	T30282	calcium-binding pr	909	6	0.4	42	2	S78170	ribosomal protein
837	7	0.5	1567	2	T03730	antigen containing	910	6	0.4	42	2	T01991	hypothetical prote
838	7	0.5	1582	2	A56248	sulfonyleurea recep	911	6	0.4	42	2	T07285	hypothetical prote
839	7	0.5	1610	2	A46227	voltage-dependent	912	6	0.4	43	2	S02031	zinc-binding prote
840	7	0.5	1611	2	G84493	probable retroelem	913	6	0.4	44	2	C34327	tropinin T, fast s
841	7	0.5	1613	2	S39059	protein BRG1 - hum	914	6	0.4	45	2	T52124	hypothetical prote
842	7	0.5	1638	2	T25352	hypothetical prote	915	6	0.4	47	2	S43651	proteasome subunit
843	7	0.5	1645	2	H85554	hypothetical prote	916	6	0.4	47	2	D40605	hmc 3'-region hypo
844	7	0.5	1646	2	JH0422	voltage-dependent	917	6	0.4	49	2	T12879	hypothetical prote
845	7	0.5	1647	2	S45252	SNF2beta protein -	918	6	0.4	50	2	B82702	hypothetical prote
846	7	0.5	1664	2	F84485	probable retroelem	919	6	0.4	51	2	T10865	hypothetical prote
847	7	0.5	1684	2	T02367	hypothetical prote	920	6	0.4	54	2	JS0853	hypothetical 5.8K
848	7	0.5	1706	2	I84499	zinc finger protei	921	6	0.4	54	2	T07284	hypothetical prote
849	7	0.5	1711	2	C71625	variant-specific s	922	6	0.4	55	2	T09464	H+-transporting AT
850	7	0.5	1717	2	T13961	DNA-directed RNA p	923	6	0.4	57	2	JN0740	hypothetical 6.7K
851	7	0.5	1720	2	T07258	cell division prot	924	6	0.4	59	2	C33556	prothymosin alpha
852	7	0.5	1721	1	I38902	retinoblastoma bin	925	6	0.4	60	1	HSAX22	histone H2B - Nile
853	7	0.5	1729	2	T18396	erythrocyte membra	926	6	0.4	61	2	S11646	hypothetical prote
854	7	0.5	1741	2	T13610	parallel sister ch	927	6	0.4	62	2	C71125	hypothetical prote
855	7	0.5	1753	2	T00350	hypothetical prote	928	6	0.4	63	1	H64095	carbon storage reg
856	7	0.5	1766	2	S03701	141K protein - pea	929	6	0.4	63	2	S07127	chymotrypsin/elast
857	7	0.5	1768	2	T13349	parallel sister ch	930	6	0.4	64	2	C64010	hypothetical prote
858	7	0.5	1774	2	B56101	collagen alpha 1(X	931	6	0.4	66	2	S61085	M protein precursor
859	7	0.5	1793	2	T47897	guanine nucleotide	932	6	0.4	66	2	H86678	hypothetical prote
860	7	0.5	1820	1	CHEE	sodium channel pro	933	6	0.4	67	2	S10555	glutathione transf
861	7	0.5	1824	2	T07589	disease resistance	934	6	0.4	69	2	A86808	hypothetical prote
862	7	0.5	1829	2	T34239	hypothetical prote	935	6	0.4	69	2	B85775	hypothetical prote
863	7	0.5	1829	2	T24583	hypothetical prote	936	6	0.4	69	2	C64925	hypothetical prote
864	7	0.5	1857	1	S01787	fatty-acid synthas	937	6	0.4	70	2	S73177	ribosomal protein
865	7	0.5	1857	2	T50513	hypothetical prote	938	6	0.4	70	2	A64011	hypothetical prote
866	7	0.5	1948	2	S00485	gene 11-1 protein	939	6	0.4	70	2	T23415	hypothetical prote
867	7	0.5	1966	2	T08991	hypothetical prote	940	6	0.4	70	2	T26824	hypothetical prote
868	7	0.5	2009	2	S49764	SEC7 protein - yea	941	6	0.4	70	2	D32870	hypothetical prote
869	7	0.5	2110	2	B41110	mycocosate synth	942	6	0.4	71	2	D42506	F-ORF-E protein -
870	7	0.5	2111	2	A70668	mycocosate synth	943	6	0.4	71	2	T32501	hypothetical prote
871	7	0.5	2149	2	C96695	ribulose biphosph	944	6	0.4	71	2	T29577	hypothetical prote
872	7	0.5	2161	2	JH0564	calcium channel al	945	6	0.4	72	2	S19237	histone H2B (clone
873	7	0.5	2181	2	A38198	calcium channel al	946	6	0.4	72	2	S78368	ribosomal protein
874	7	0.5	2203	2	T42742	voltage-dependent	947	6	0.4	72	2	C47072	marB protein - Esc
875	7	0.5	2206	2	G71611	hypothetical prote	948	6	0.4	72	2	B42645	ribosomal protein
876	7	0.5	2225	2	T26063	hypothetical prote	949	6	0.4	72	2	T47513	hypothetical prote
877	7	0.5	2265	2	T26183	hypothetical prote	950	6	0.4	72	2	G81564	ribosomal protein
878	7	0.5	2282	2	T42717	DNA-binding protei	951	6	0.4	73	2	S72246	dynelin heavy chain
879	7	0.5	2457	2	T18492	hypothetical prote	952	6	0.4	73	2	H71149	hypothetical prote
880	7	0.5	2924	2	T18378	variant-specific s	953	6	0.4	74	2	A05059	photosystem II pho
881	7	0.5	3026	2	T28431	variant surface pr	954	6	0.4	75	2	T03564	ribosomal protein
882	7	0.5	3135	2	A48584	transmission block	955	6	0.4	75	2	JQ2358	wheat aluminum ind
883	7	0.5	3147	2	T18674	hypothetical prote	956	6	0.4	76	2	S22203	photosystem I chal
884	7	0.5	3418	1	G02334	breast cancer tumo	957	6	0.4	76	2	S09897	hypothetical prote
885	7	0.5	4006	2	T29042	probable tenascin	958	6	0.4	76	2	T18091	hypothetical prote
886	7	0.5	4385	2	T29042	hypothetical prote	959	6	0.4	77	2	A47070	ferrochelatase (EC
887	7	0.5	4845	2	T31067	BIR repeat contain	960	6	0.4	78	2	C49039	T-cell receptor be
888	7	0.5	4859	2	S74173	ryanodine receptor	961	6	0.4	79	2	S76591	hypothetical prote
889	7	0.5	4859	2	S66572	ryanodine receptor	962	6	0.4	79	2	E64354	conserved hypoteth
890	7	0.5	4872	2	S27272	ryanodine receptor	963	6	0.4	79	2	A86517	hypothetical prote
891	7	0.5	5035	1	I46646	ryanodine receptor	964	6	0.4	79	2	T35842	hypothetical prote
892	7	0.5	5037	1	A54161	ryanodine-binding	965	6	0.4	79	2	B72106	hypothetical prote
893	7	0.5	5037	2	B35041	ryanodine receptor	966	6	0.4	79	2	T13293	hypothetical prote
894	7	0.5	5170	2	T15348	hypothetical prote	967	6	0.4	79	2	C84077	nitrogen fixation
895	7	0.5	5175	2	T20992	hypothetical prote	968	6	0.4	81	2	T31007	conserved hypoteth
896	7	0.5	5198	2	T43290	hemocentin precurs	969	6	0.4	81	2	D81565	hypothetical prote
897	6	0.4	12	2	S16335	beta-conglycinin a	970	6	0.4	81	2	F82745	hypothetical prote
898	6	0.4	18	2	A59137	protein P11 - gold	971	6	0.4	83	2	I46058	caldesmon - bovine
899	6	0.4	20	2	S19240	histone H2B (clone	972	6	0.4	83	2	H69175	hypothetical prote
900	6	0.4	20	2	A61093	glue protein - Cal	973	6	0.4	83	2	A57231	caldesmon - leech
901	6	0.4	22	2	B46236	transforming prote	974	6	0.4	84	2	T10345	hypothetical prote
902	6	0.4	22	2	S69351	N-methylhydantoin	975	6	0.4	87	2	S26963	hypothetical prote
903	6	0.4	30	2	PC2328	multicatalytic end	976	6	0.4	87	2	D82136	cell division topo
904	6	0.4	30	2	PD0014	cAMP response elem	977	6	0.4	87	2	JQ1683	UL49A protein - hu
905	6	0.4	37	2	G82917	ribosomal protein	978	6	0.4	88	1	QOEC8	hypothetical prote

687	7	0.5	886	2	A48586	suppressor of hair	760	7	0.5	1114	2	I50222	deltaErf1 - chicken
688	7	0.5	886	2	A54442	3',5'-cyclic-nucle	761	7	0.5	1115	2	B84476	probable TPR repea
689	7	0.5	888	2	T51593	GTP-binding regula	762	7	0.5	1116	2	T13854	nuclear protein SA
690	7	0.5	890	2	T00800	disease resistance	763	7	0.5	1117	2	JC4934	delta-crystallin/E
691	7	0.5	892	2	B46203	mating type A alph	764	7	0.5	1121	2	T25715	hypothetical prote
692	7	0.5	895	2	A45554	rhostry complex po	765	7	0.5	1124	2	JX0293	zinc finger protei
693	7	0.5	898	2	H84701	probable villin li	766	7	0.5	1137	2	S13759	morphogenesis-rela
694	7	0.5	899	2	F84477	probable retroelem	767	7	0.5	1141	2	A44093	CGMP-inhibited cAM
695	7	0.5	899	2	B48586	suppressor of hair	768	7	0.5	1148	2	A71446	mypl protein - smu
696	7	0.5	901	2	T01135	probable GTP-bind	769	7	0.5	1150	2	S98775	hypothetical prote
697	7	0.5	901	2	JC6093	dead ringer nuclea	770	7	0.5	1153	2	T00615	DNA polymerase III
698	7	0.5	902	2	G83635	probable clpA/B-ty	771	7	0.5	1154	2	A56242	E-box-binding repr
699	7	0.5	906	2	T01440	hypothetical prote	772	7	0.5	1161	2	I59311	narolysin (EC 3.4
700	7	0.5	907	2	T48500	Me12-like protein	773	7	0.5	1167	2	T13927	adenylate cyclase
701	7	0.5	909	2	S32538	CGMP-gated cation	774	7	0.5	1169	2	C84547	hypothetical prote
702	7	0.5	910	2	T50671	villin 1 (imported	775	7	0.5	1171	2	T13065	PIP82 protein - fr
703	7	0.5	912	2	A54423	brevican precursor	776	7	0.5	1181	2	A43346	1-phosphatidylinos
704	7	0.5	913	2	T15474	hypothetical prote	777	7	0.5	1182	2	T30189	myelin transcripti
705	7	0.5	929	2	T28927	hypothetical prote	778	7	0.5	1184	2	A42904	adenylcyclase ly
706	7	0.5	932	2	S62555	protoptast regene	779	7	0.5	1184	2	T41515	coiled coil protei
707	7	0.5	938	2	A56731	chromatin assembly	780	7	0.5	1187	2	T46637	transcription fact
708	7	0.5	943	2	B83068	probable oxidoredu	781	7	0.5	1188	2	T05324	hypothetical prote
709	7	0.5	944	2	G85720	translation initia	782	7	0.5	1188	2	T46608	zinc finger protei
710	7	0.5	946	2	T19465	hypothetical prote	783	7	0.5	1196	2	T13057	KIAA0729 protein -
711	7	0.5	949	2	D84487	probable retroelem	784	7	0.5	1197	2	T13956	timeless protein h
712	7	0.5	954	2	T24748	hypothetical prote	785	7	0.5	1197	2	D86317	protein F15H18.21
713	7	0.5	954	2	T22369	hypothetical prote	786	7	0.5	1199	2	T15826	hypothetical prote
714	7	0.5	957	2	G84528	hypothetical prote	787	7	0.5	1199	2	S20969	Na+/Ca2+, K+-exchan
715	7	0.5	960	2	T07680	VPS41 protein homo	788	7	0.5	1205	2	T13959	timeless protein T
716	7	0.5	960	2	T17297	hypothetical prote	789	7	0.5	1208	2	T00362	hypothetical prote
717	7	0.5	963	2	A5926	DNA binding protei	790	7	0.5	1210	2	A48001	phospholipase C (E
718	7	0.5	966	2	E84053	penicillin-binding	791	7	0.5	1214	2	T00356	hypothetical prote
719	7	0.5	967	1	HXAD2	hexon protein - hu	792	7	0.5	1217	2	T00270	hypothetical prote
720	7	0.5	968	2	T45746	hypothetical prote	793	7	0.5	1219	2	I61713	co-repressor prote
721	7	0.5	971	2	S23408	prematurely termin	794	7	0.5	1221	2	A49457	fibulin-2 precurs
722	7	0.5	972	2	T50400	related to actin-i	795	7	0.5	1223	2	S29717	adenylate cyclase
723	7	0.5	972	2	T49773	hypothetical prote	796	7	0.5	1227	2	B34911	band 3-related pro
724	7	0.5	973	2	T01862	hypothetical prote	797	7	0.5	1227	2	A33638	erythrocyte anion
725	7	0.5	975	1	A31497	kinesin heavy chai	798	7	0.5	1229	2	A56068	co-repressor prote
726	7	0.5	978	2	H86319	hypothetical prote	799	7	0.5	1231	1	A48490	endo-1,4-beta-xyla
727	7	0.5	979	2	A35913	regulatory factor	800	7	0.5	1232	2	I38496	anion exchanger 3
728	7	0.5	993	2	T17230	hypothetical prote	801	7	0.5	1232	2	T47993	hypothetical prote
729	7	0.5	999	2	S68689	glucose regulated	802	7	0.5	1233	2	S56271	hypothetical prote
730	7	0.5	1001	2	S30385	G9a protein - huma	803	7	0.5	1241	2	J00466	potassium transpor
731	7	0.5	1013	2	B96544	hypothetical prote	804	7	0.5	1262	2	T33074	hypothetical prote
732	7	0.5	1014	2	T31433	Na+/Ca2+, K+-exchan	805	7	0.5	1264	2	S41603	type V adenylyl cy
733	7	0.5	1022	2	I53078	homeotic gene regu	806	7	0.5	1268	2	T50252	probable transcrip
734	7	0.5	1026	2	B96663	hypothetical prote	807	7	0.5	1272	2	C96637	hypothetical prote
735	7	0.5	1032	2	T18293	quanylate kinase-i	808	7	0.5	1274	2	T04018	hypothetical prote
736	7	0.5	1034	2	S76134	hypothetical prote	809	7	0.5	1277	2	T32731	PAR interacting pr
737	7	0.5	1037	2	A60163	glycoprotein Iib -	810	7	0.5	1277	2	T14152	synaptic scaffold
738	7	0.5	1038	2	T02634	rep protein homolo	811	7	0.5	1280	2	T00365	hypothetical prote
739	7	0.5	1038	2	S52522	hypothetical prote	812	7	0.5	1283	2	T28812	hypothetical prote
740	7	0.5	1043	2	A56037	DNA-binding protei	813	7	0.5	1300	2	T03166	probable immediate
741	7	0.5	1045	2	T41119	internalin- relate	814	7	0.5	1360	2	T06699	zinc finger protei
742	7	0.5	1047	2	T21306	hypothetical prote	815	7	0.5	1361	2	S50943	hypothetical prote
743	7	0.5	1054	2	T14189	hypothetical prote	816	7	0.5	1386	2	T49316	profilaggrin relat
744	7	0.5	1058	2	T30580	p-type ATPase - sl	817	7	0.5	1388	2	T38720	chromodomain helic
745	7	0.5	1062	2	S61196	SUM1 protein - yea	818	7	0.5	1389	2	I50090	carboxypeptidase g
746	7	0.5	1063	2	D86731	hypothetical prote	819	7	0.5	1402	2	I46707	translation initia
747	7	0.5	1067	2	T28663	hypothetical prote	820	7	0.5	1428	2	T13926	probable protein p
748	7	0.5	1067	2	S35423	protein kinase sgg	821	7	0.5	1432	2	S58819	antiviral protein -
749	7	0.5	1068	2	G86452	protein AAR3127.1	822	7	0.5	1440	2	T27942	lin-15B protein -
750	7	0.5	1072	2	T50949	verprolin related	823	7	0.5	1443	2	I50600	neogenin - chicken
751	7	0.5	1075	2	T49879	hypothetical prote	824	7	0.5	1463	2	C86482	protein F5J5.1 [im
752	7	0.5	1078	2	T42712	myelin transcripti	825	7	0.5	1467	2	T48162	hypothetical prote
753	7	0.5	1087	2	T30330	geisolin-related p	826	7	0.5	1474	2	T20488	hypothetical prote
754	7	0.5	1094	2	S49313	protein kinase - s	827	7	0.5	1474	2	F69009	probable membrane
755	7	0.5	1097	2	S68685	adenylate cyclase	828	7	0.5	1479	2	T17401	transcription regu
756	7	0.5	1099	2	A55405	adenylate cyclase	829	7	0.5	1480	2	T05566	hypothetical prote
757	7	0.5	1100	2	T42260	quanylate cyclase	830	7	0.5	1493	2	A44224	DNA repair helicas
758	7	0.5	1108	2	A48508	cyclic-nucleotide	831	7	0.5	1511	2	T42711	sulfonylurea recep
759	7	0.5	1110	2	I51116	NF-180 - sea lamp	832	7	0.5	1545	2	T42751	sulfonylurea recep

541	7	0.5	590	2	S57594	hypothetical prote	614	7	0.5	687	2	C57713	chloride channel C
542	7	0.5	591	2	S51303	hypothetical prote	615	7	0.5	692	2	S37976	hypothetical prote
543	7	0.5	592	2	B48315	lamin B2 - mouse	616	7	0.5	694	2	T52574	cyclic nucleotide
544	7	0.5	593	2	I51213	drebrin - chicken	617	7	0.5	699	2	A34660	histidine rich cal
545	7	0.5	595	2	A43534	Lupus autoantigen	618	7	0.5	700	2	T05841	spliceosome-assoc1
546	7	0.5	596	2	T23193	hypothetical prote	619	7	0.5	707	1	DNMS	nucleolin - mouse
547	7	0.5	599	2	A45195	adenylcyclase ty	620	7	0.5	707	2	A35804	nucleolin - human
548	7	0.5	596	2	T18316	hypothetical prote	621	7	0.5	708	2	S53411	hypothetical prote
549	7	0.5	599	2	S18735	centromere protein	622	7	0.5	713	2	A27441	nucleolin - Chines
550	7	0.5	602	2	A36715	exo-poly-alpha-gal	623	7	0.5	715	2	D85087	hypothetical prote
551	7	0.5	603	2	T11284	NADH dehydrogenase	624	7	0.5	719	2	S51739	transcription repr
552	7	0.5	603	2	E84744	hypothetical prote	625	7	0.5	720	2	T43327	gluconate transpor
553	7	0.5	606	2	S57552	hypothetical prote	626	7	0.5	721	2	S29795	hypothetical prote
554	7	0.5	606	2	S70358	centromere protein	627	7	0.5	725	2	T42688	hypothetical prote
555	7	0.5	607	2	A43776	drebrin E2 - chick	628	7	0.5	728	2	A54603	transcription fact
556	7	0.5	608	2	T06632	hypothetical prote	629	7	0.5	736	2	T03849	Fas-binding protei
557	7	0.5	610	2	S67701	hypothetical prote	630	7	0.5	740	2	T03847	Fas-binding protei
558	7	0.5	610	2	A96701	protein F12A21.3 [631	7	0.5	741	2	T40095	zinc finger, C3HC4
559	7	0.5	617	2	S42719	actin-binding prot	632	7	0.5	751	2	A49974	beta-amyloid precu
560	7	0.5	618	1	PRECT4	proteinase IV (EC	633	7	0.5	752	2	S64750	probable Arp-depen
561	7	0.5	618	1	S38004	probable transport	634	7	0.5	757	2	T50959	hypothetical prote
562	7	0.5	618	2	D85786	proteinase IV, a s	635	7	0.5	759	2	S25330	proteol protein - yea
563	7	0.5	618	2	T52076	AB13-interacting p	636	7	0.5	760	2	G86373	protein T23E23.14
564	7	0.5	618	2	T52075	pseudo-response re	637	7	0.5	763	2	A49321	amyloid beta (A4)
565	7	0.5	619	2	E84800	hypothetical prote	638	7	0.5	763	2	S51300	probable membrane
566	7	0.5	620	2	JH0821	95K golgi antigen	639	7	0.5	765	2	S42880	amyloid precursor-
567	7	0.5	622	2	F71342	calnexin precursor	640	7	0.5	765	2	S22314	transcription fact
568	7	0.5	625	2	F70459	conserved hypothet	641	7	0.5	767	2	S41479	DNA-binding protei
569	7	0.5	626	2	T08926	hypothetical prote	642	7	0.5	768	2	H54024	protein kinase (EC
570	7	0.5	627	2	H86180	hypothetical prote	643	7	0.5	771	2	T02565	disease resistance
571	7	0.5	628	2	G71565	probable signal pe	644	7	0.5	772	2	I50463	protein kinase - c
572	7	0.5	629	2	T38214	hypothetical prote	645	7	0.5	776	2	B96666	protein F22C12.6 [
573	7	0.5	629	2	T48799	Rrp9p related prot	646	7	0.5	777	2	B54024	protein kinase (EC
574	7	0.5	630	2	S29796	hypothetical prote	647	7	0.5	777	2	F54024	protein kinase (EC
575	7	0.5	630	2	T48369	hypothetical prote	648	7	0.5	777	2	T08659	ral guanine nucleo
576	7	0.5	631	2	G96701	unknown protein, 7	649	7	0.5	779	2	E54024	protein kinase (EC
577	7	0.5	632	2	T46504	hypothetical prote	650	7	0.5	780	2	F96840	hypothetical prote
578	7	0.5	633	2	T14612	hypothetical prote	651	7	0.5	780	2	I47038	hypothetical prote
579	7	0.5	635	2	H69626	PTS fructose-speci	652	7	0.5	783	2	A55817	vasopressin-activa
580	7	0.5	637	2	JH0674	L-proline transpor	653	7	0.5	784	2	PN0009	cyclin-dependent k
581	7	0.5	639	2	T46577	arylsulfatase (EC	654	7	0.5	786	2	A35466	neurofilament trip
582	7	0.5	642	2	T23687	hypothetical prote	655	7	0.5	788	2	C70441	progesterone recep
583	7	0.5	644	2	S55395	neurofilament prot	656	7	0.5	789	2	T51310	lipote-protein 11
584	7	0.5	645	1	A23723	protein disulfide-	657	7	0.5	789	2	T52067	RNA helicase RH28
585	7	0.5	646	2	S55048	protein-tyrosine k	658	7	0.5	792	2	S63141	hypothetical prote
586	7	0.5	648	2	G85056	probable receptor-	659	7	0.5	792	2	T42963	hypothetical prote
587	7	0.5	648	2	F69848	transcription anti	660	7	0.5	794	2	T27870	hypothetical prote
588	7	0.5	649	2	T39826	probable clathrin	661	7	0.5	798	2	D96563	probable bZIP prot
589	7	0.5	651	2	T15624	hypothetical prote	662	7	0.5	798	2	T33022	hypothetical prote
590	7	0.5	653	2	A39922	potassium channel	663	7	0.5	798	2	JC7500	glik protein - chic
591	7	0.5	654	2	S11049	potassium channel	664	7	0.5	799	2	JH0797	castor protein - f
592	7	0.5	660	2	B96555	hypothetical prote	665	7	0.5	802	1	S48529	NAB3 protein - yea
593	7	0.5	661	2	T24584	hypothetical prote	666	7	0.5	806	2	S22765	heterogeneous ribo
594	7	0.5	661	2	A81125	bacteriophage tran	667	7	0.5	808	2	T23129	hypothetical prote
595	7	0.5	664	2	C70584	probable serine-th	668	7	0.5	809	1	JQ0032	anthrax toxin leth
596	7	0.5	668	2	T05257	probable disease r	669	7	0.5	809	2	S67665	ubiquitin-specific
597	7	0.5	669	2	S78050	high mobility grou	670	7	0.5	817	2	JC4176	pyruvate,water dik
598	7	0.5	672	2	T32557	hypothetical prote	671	7	0.5	819	1	S40400	protein kinase SWE
599	7	0.5	674	2	F81990	NADH dehydrogenase	672	7	0.5	819	2	F75196	pyruvate,water dik
600	7	0.5	674	2	D81220	NADH dehydrogenase	673	7	0.5	820	2	T00645	hypothetical prote
601	7	0.5	674	2	D82118	conserved hypothet	674	7	0.5	821	2	B71229	pyruvate,water dik
602	7	0.5	675	2	T03744	myoD protein inhib	675	7	0.5	825	2	S54465	YTA12 protein prec
603	7	0.5	676	2	T01084	hypothetical prote	676	7	0.5	830	2	T07743	probable 1,4-alpha
604	7	0.5	678	2	A54514	glutamic acid-rich	677	7	0.5	845	2	JC5256	adipocyte transcri
605	7	0.5	678	2	T43539	spindle checkpoint	678	7	0.5	852	2	A34373	histidine-rich cal
606	7	0.5	679	2	S48939	hypothetical prote	679	7	0.5	855	2	T10665	hypothetical prote
607	7	0.5	679	2	A83624	probable chemotaxi	680	7	0.5	858	2	T08881	prominin - mouse
608	7	0.5	680	2	T19768	hypothetical prote	681	7	0.5	859	2	C96504	protein F9C16.23 [
609	7	0.5	683	2	D71433	hypothetical prote	682	7	0.5	862	2	T34342	hypothetical prote
610	7	0.5	684	2	T51410	N2,N2-dimethylguan	683	7	0.5	883	2	A71434	probable RNA helic
611	7	0.5	686	2	A45483	chloride channel,	684	7	0.5	884	2	S66308	nitrate reductase
612	7	0.5	687	2	A57713	chloride channel C	685	7	0.5	886	1	RGBYL3	regulatory protein
613	7	0.5	687	2	B57713	chloride channel K	686	7	0.5	886	2	A59223	nitrate reductase

395	7	0.5	427	2	S38002	hypothetical prote	468	7	0.5	495	2	S10851	glycinin G1 precu
396	7	0.5	429	2	I49603	transcription regu	469	7	0.5	495	2	C71410	hypothetical prote
397	7	0.5	429	2	F86240	hypothetical prote	470	7	0.5	496	2	S68160	probable RNA bindi
398	7	0.5	430	2	I57013	guanine nucleotide	471	7	0.5	497	2	JC5076	myc-associated zin
399	7	0.5	431	2	C69087	dihydrolipoamide d	472	7	0.5	498	2	T48385	transporter like p
400	7	0.5	433	2	JC4529	guanylate cyclase	473	7	0.5	500	2	S26688	legumin K - garden
401	7	0.5	434	2	D75352	deoxyguanosinetrip	474	7	0.5	500	2	T49017	hypothetical prote
402	7	0.5	434	2	C71418	cytochrome P450 -	475	7	0.5	502	2	I52637	Ca2+/calmodulin-de
403	7	0.5	435	2	S31290	cyclin B5 - yeast	476	7	0.5	502	2	S01401	H+-transporting AT
404	7	0.5	435	2	T01826	microfibril-associ	477	7	0.5	502	2	F31482	H+-transporting AT
405	7	0.5	436	2	T38812	hypothetical prote	478	7	0.5	502	2	T48560	hypothetical prote
406	7	0.5	440	2	S71795	transcription fact	479	7	0.5	502	2	T01179	hypothetical prote
407	7	0.5	445	2	S43328	tubulin beta-7 cha	480	7	0.5	503	2	S00336	legumin B leagu pre
408	7	0.5	445	2	T47813	hypothetical prote	481	7	0.5	504	1	A32965	cytochrome P450 4A
409	7	0.5	446	2	A32284	chromogranin A pre	482	7	0.5	504	2	I53868	alpha-internexin -
410	7	0.5	448	2	T06698	hypothetical prote	483	7	0.5	504	2	G02474	interferon regulat
411	7	0.5	448	2	S05355	hypothetical prote	484	7	0.5	504	2	T10558	hypothetical prote
412	7	0.5	449	2	J40049	Tubulin beta-2 cha	485	7	0.5	505	2	A41023	alpha-internexin -
413	7	0.5	449	2	A5709	major early-transc	486	7	0.5	507	1	A32966	cytochrome P450 4A
414	7	0.5	450	2	A37223	alpha-2B-adrenergi	487	7	0.5	507	2	S18872	legumin-like stora
415	7	0.5	450	2	T10632	hypothetical prote	488	7	0.5	509	2	E86427	hypothetical prote
416	7	0.5	452	2	T20556	hypothetical prote	489	7	0.5	511	2	A40110	nicotinic acetylch
417	7	0.5	452	2	I39383	angio-associated m	490	7	0.5	511	2	JC1404	CDEI-box DNA-bindi
418	7	0.5	454	2	H86231	hypothetical prote	491	7	0.5	516	2	T48394	hypothetical prote
419	7	0.5	455	1	B25493	indoleacetamide hy	492	7	0.5	517	2	T23499	hypothetical prote
420	7	0.5	455	2	S65157	hypothetical prote	493	7	0.5	517	2	A49413	perilipin A - rat
421	7	0.5	455	2	I48724	zinc finger protei	494	7	0.5	518	2	T24783	hypothetical prote
422	7	0.5	456	2	D84651	hypothetical prote	495	7	0.5	522	2	JC4024	poliovirus recepto
423	7	0.5	457	1	A28468	chromogranin A pre	496	7	0.5	522	2	A55194	importin 1 - Afric
424	7	0.5	461	2	B93601	probable transport	497	7	0.5	523	2	B55194	importin 2 - Afric
425	7	0.5	462	2	A40552	bindin fertilizati	498	7	0.5	523	2	S67271	hypothetical prote
426	7	0.5	463	2	S28469	phosphomannomutase	499	7	0.5	524	2	A86356	hypothetical prote
427	7	0.5	464	2	S45363	LEO1 protein - yea	500	7	0.5	526	2	G83436	hypothetical prote
428	7	0.5	466	2	T22141	hypothetical prote	501	7	0.5	527	1	A71418	cytochrome P450 d1
429	7	0.5	467	2	G84937	glutamate--tRNA li	502	7	0.5	527	2	A32469	80K protein H prec
430	7	0.5	467	2	S61105	hypothetical prote	503	7	0.5	528	1	ACCH2N	nicotinic acetylch
431	7	0.5	470	2	S47887	BFR1 protein - yea	504	7	0.5	531	2	A55887	caldesmon, non-mus
432	7	0.5	471	2	B83127	probable secretion	505	7	0.5	531	2	T33980	hypothetical prote
433	7	0.5	471	2	T49019	probable RNA bindi	506	7	0.5	531	2	G83268	conserved hypothet
434	7	0.5	473	2	S22321	deoxyribodipyrimid	507	7	0.5	531	2	T49058	hypothetical prote
435	7	0.5	474	1	TVR7C4	Ca2+/calmodulin-de	508	7	0.5	533	2	T34458	hypothetical prote
436	7	0.5	474	2	S28419	lamin B-3 - mouse	509	7	0.5	534	2	S61185	hypothetical prote
437	7	0.5	475	2	A71110	hypothetical prote	510	7	0.5	538	2	T03262	cytochrome P450 -
438	7	0.5	475	2	S28933	hypothetical prote	511	7	0.5	541	2	T48811	hypothetical prote
439	7	0.5	475	2	C68683	N-acetylmuramoyl-L	512	7	0.5	544	2	A57597	beta-p21-activated
440	7	0.5	477	2	F85169	cytochrome P450 li	513	7	0.5	544	2	I49376	p21 activated kina
441	7	0.5	477	2	A72736	zinc-finger protei	514	7	0.5	545	2	H86322	calcium-dependent
442	7	0.5	478	2	I53960	PRR2 alpha - human	515	7	0.5	545	2	S68687	thermosome - Metha
443	7	0.5	478	2	S69004	hypothetical prote	516	7	0.5	547	2	T45635	hypothetical prote
444	7	0.5	480	4	TVHUBD	transforming prote	517	7	0.5	550	2	B84900	hypothetical prote
445	7	0.5	480	2	S14394	bindin precursor -	518	7	0.5	552	2	S45886	hypothetical prote
446	7	0.5	481	2	T18465	hypothetical prote	519	7	0.5	552	2	H70445	hypothetical prote
447	7	0.5	481	2	A26483	bindin precursor -	520	7	0.5	553	2	T06499	hypothetical prote
448	7	0.5	482	1	XUBOLA	dihydrolipoamide S	521	7	0.5	554	2	A56730	carl protein - Pod
449	7	0.5	482	2	A32422	dihydrolipoamide S	522	7	0.5	559	1	WMBES1	71k upper matrix p
450	7	0.5	482	2	S65760	dihydrolipoamide t	523	7	0.5	559	2	T48180	hypothetical prote
451	7	0.5	483	2	S37845	transcription init	524	7	0.5	560	2	T02404	probable beta-gluc
452	7	0.5	483	2	S27880	Nasopressin recept	525	7	0.5	561	2	JC2436	5'-nucleotidase (E
453	7	0.5	484	2	A24942	legumin B4 precurs	526	7	0.5	566	2	T06453	probable legumin B
454	7	0.5	484	2	T06063	hypothetical prote	527	7	0.5	566	2	S28068	mxid protein - Shi
455	7	0.5	484	2	S61870	hypothetical prote	528	7	0.5	569	2	S56658	stress-induced pro
456	7	0.5	485	2	B40552	bindin fertilizati	529	7	0.5	570	1	B38546	major structural n
457	7	0.5	486	2	JC7241	calcineurin cataly	530	7	0.5	572	2	T22547	hypothetical prote
458	7	0.5	486	2	C96649	hypothetical prote	531	7	0.5	580	2	T43485	hypothetical prote
459	7	0.5	486	2	S67589	probable membrane	532	7	0.5	580	2	G86169	hypothetical prote
460	7	0.5	491	2	H84477	probable PttA-like	533	7	0.5	581	2	T46024	hypothetical prote
461	7	0.5	491	2	S49790	probable membrane	534	7	0.5	581	2	C96538	hypothetical prote
462	7	0.5	492	2	A61382	phosphorylation re	535	7	0.5	582	2	S53814	DEAD box protein -
463	7	0.5	493	1	S46283	calcium-dependent	536	7	0.5	582	2	S19424	hypothetical prote
464	7	0.5	493	2	A26579	cytochrome P450 2E	537	7	0.5	584	2	T23174	hypothetical prote
465	7	0.5	493	2	G84594	probable diacylgly	538	7	0.5	586	2	C30411	synapsin Iia - rat
466	7	0.5	494	2	A42170	zinc finger protei	539	7	0.5	587	2	B83015	arginyl-tRNA synth
467	7	0.5	495	1	FWSYG2	glycinin chain Ala	540	7	0.5	587	2	JC5300	Ran GTPase activat

249	7	0.5	280	2	S35103	bone sialoprotein	322	7	0.5	340	2	G82047	conserved hypothet
250	7	0.5	280	2	H72389	transcription regu	323	7	0.5	340	2	T51386	probable protein w
251	7	0.5	282	1	D64757	ykG protein - Esc	324	7	0.5	341	2	T41450	hypothetical prote
252	7	0.5	284	2	B48516	surfactant protein	325	7	0.5	343	2	S45321	folliculin - mous
253	7	0.5	288	2	I51332	signal sequence re	326	7	0.5	343	2	T02399	hypothetical prote
254	7	0.5	288	2	T32459	hypothetical prote	327	7	0.5	344	2	I57698	folliculin - rat
255	7	0.5	289	2	JC7279	Down syndrome crit	328	7	0.5	347	2	T33204	hypothetical prote
256	7	0.5	292	2	E72223	conserved hypothet	329	7	0.5	347	2	S40105	gene AGP2beta-2 pr
257	7	0.5	292	2	A84427	hypothetical prote	330	7	0.5	350	2	S00337	legumin B Legk pre
258	7	0.5	292	2	T00996	En/Spm-like transp	331	7	0.5	350	2	T06107	probable serine/th
259	7	0.5	293	2	T09170	ribosomal protein	332	7	0.5	350	2	T33458	hypothetical prote
260	7	0.5	295	2	A38977	cyclin D1 - human	333	7	0.5	351	2	F96597	hypothetical prote
261	7	0.5	295	2	T39012	hypothetical prote	334	7	0.5	353	2	A86438	hypothetical prote
262	7	0.5	295	2	E84862	hypothetical prote	335	7	0.5	354	2	I48722	zinc finger protei
263	7	0.5	296	2	G86523	hypothetical prote	336	7	0.5	357	2	JC4703	basic helix-loop-h
264	7	0.5	296	2	G72099	conserved hypothet	337	7	0.5	357	2	I49338	neurogenic differe
265	7	0.5	298	2	A71159	probable membrane	338	7	0.5	357	2	D71262	probable hemolysin
266	7	0.5	298	2	C86871	conserved hypothet	339	7	0.5	359	2	T21840	hypothetical prote
267	7	0.5	299	2	F75003	stomatol-like prot	340	7	0.5	363	2	D64640	hypothetical prote
268	7	0.5	299	2	A54090	pxF protein - Chin	341	7	0.5	366	2	A96692	hypothetical prote
269	7	0.5	299	2	I37468	house keeping gene	342	7	0.5	368	1	TVMSML	transforming prote
270	7	0.5	299	2	H84661	heme oxygenase 2 (343	7	0.5	368	2	H96712	probable DNA-bindi
271	7	0.5	299	2	T02515	cytoskeletal prote	344	7	0.5	369	2	D81984	riboflavin bifunct
272	7	0.5	301	2	T20651	hypothetical prote	345	7	0.5	371	2	T21707	hypothetical prote
273	7	0.5	301	2	JW0079	heterogeneous nucl	346	7	0.5	372	2	T02550	NPk1-related prote
274	7	0.5	302	2	E86267	hypothetical prote	347	7	0.5	372	2	S76427	hypothetical prote
275	7	0.5	303	1	S28392	protein-tyrosine-p	348	7	0.5	375	2	A64348	hypothetical prote
276	7	0.5	303	2	S12867	carbonate dehydrat	349	7	0.5	375	2	S26059	probable transform
277	7	0.5	303	2	T10653	hypothetical prote	350	7	0.5	376	2	T34734	hypothetical prote
278	7	0.5	303	2	T48493	hypothetical prote	351	7	0.5	376	2	T10455	heat shock related
279	7	0.5	304	2	S44897	ZK1236.2 protein -	352	7	0.5	381	2	G02668	neurogenic basic-h
280	7	0.5	306	2	G70481	thiamin monophosph	353	7	0.5	381	2	JC4647	KW8 protein - rat
281	7	0.5	306	2	T32461	hypothetical prote	354	7	0.5	381	2	T49544	hypothetical prote
282	7	0.5	310	2	I46987	bone sialoprotein	355	7	0.5	382	2	T11219	ubiquinol--cytochr
283	7	0.5	312	2	S19378	hypothetical prote	356	7	0.5	382	2	S47747	2-dehydro-3-deoxyr
284	7	0.5	312	2	S66952	hypothetical prote	357	7	0.5	382	2	S72169	DNA-binding protei
285	7	0.5	314	2	T16300	hypothetical prote	358	7	0.5	382	2	G86025	ketodeoxygluconoki
286	7	0.5	316	2	T11021	farnesyltransferts	359	7	0.5	383	1	A48222	dematin 48k chain
287	7	0.5	316	2	G96513	hypothetical prote	360	7	0.5	383	2	JC4688	neuro D-related fa
288	7	0.5	316	2	S73683	MG338 homolog P02-	361	7	0.5	386	2	S41497	thyroid hormone re
289	7	0.5	317	1	GEH093	bone sialoprotein	362	7	0.5	387	2	A83393	probable acyl-CoA
290	7	0.5	317	2	S67493	huntingtin-associa	363	7	0.5	387	2	S02708	troponin T - fruit
291	7	0.5	320	2	T49591	probable 35 kDa ri	364	7	0.5	388	2	T00641	hypothetical prote
292	7	0.5	320	1	GERTS	bone sialoprotein	365	7	0.5	388	3	JC7510	benzoate X recepto
293	7	0.5	320	2	S72555	huntingtin-associa	366	7	0.5	392	1	S55971	probable peptidylp
294	7	0.5	323	2	F70646	probable quinone o	367	7	0.5	393	2	T43401	transcription init
295	7	0.5	324	2	I49768	bone sialoprotein	368	7	0.5	394	2	S64505	HGH1 protein - yea
296	7	0.5	324	2	D72070	conserved hypothet	369	7	0.5	394	2	F75006	hypothetical prote
297	7	0.5	324	2	B86555	CT429 hypothetical	370	7	0.5	396	2	S13251	troponin T - fruit
298	7	0.5	327	2	G33282	DNA-binding protei	371	7	0.5	400	2	A26258	endoplasmin - gold
299	7	0.5	328	2	S71425	DNA-directed RNA p	372	7	0.5	400	2	A55647	phyllorod - fruit
300	7	0.5	328	2	B84545	hypothetical prote	373	7	0.5	405	1	JH0795	calreticulin precu
301	7	0.5	329	1	RNECA	DNA-directed RNA p	374	7	0.5	405	2	I39062	dematin 52K chain
302	7	0.5	329	1	A41658	DNA-directed RNA p	375	7	0.5	406	2	A48059	oncoprotein zc-Myc
303	7	0.5	329	1	E72068	carbohydrate isome	376	7	0.5	406	2	S59948	aminomethyltransf
304	7	0.5	329	2	H84987	DNA-directed RNA p	377	7	0.5	409	2	E86336	hypothetical prote
305	7	0.5	329	2	D85994	RNA polymerase, al	378	7	0.5	411	2	S48647	peptidylprolyl iso
306	7	0.5	329	2	S07577	legumin storage pr	379	7	0.5	412	2	A55320	immunophilin FRB4
307	7	0.5	329	2	B86556	GutQ/KpsF family s	380	7	0.5	412	2	B86391	hypothetical prote
308	7	0.5	329	2	E71516	hypothetical prote	381	7	0.5	412	2	G02453	NN8-4AG - human (f
309	7	0.5	331	2	T41507	hypothetical prote	382	7	0.5	414	2	T49459	hypothetical prote
310	7	0.5	331	2	B47236	zinc-finger protei	383	7	0.5	415	2	JC7167	C kinase 1 interac
311	7	0.5	332	2	A81675	conserved hypothet	384	7	0.5	416	1	A42879	advanced glycosyla
312	7	0.5	332	2	E86448	hypothetical prote	385	7	0.5	416	2	T10623	hypothetical prote
313	7	0.5	333	2	D83113	DNA-directed RNA p	386	7	0.5	417	2	H72362	3-isopropylmalate
314	7	0.5	335	2	S07576	legumin storage pr	387	7	0.5	417	2	A49129	neuroblast cell li
315	7	0.5	335	2	B86429	hypothetical prote	388	7	0.5	420	2	B86217	protein T27G7.4 [i
316	7	0.5	336	2	G69091	ribosomal protein	389	7	0.5	420	2	T46910	hypothetical prote
317	7	0.5	336	2	F71942	holliday junction	390	7	0.5	422	2	T41888	IE-2 (IE-N) orf151
318	7	0.5	336	2	G64652	DNA-binding protei	391	7	0.5	423	2	B86214	hypothetical prote
319	7	0.5	337	2	S31131	hypothetical prote	392	7	0.5	424	2	T01383	GTPase-activating
320	7	0.5	338	2	S04321	legumin B (clone p	393	7	0.5	424	2	S43560	coiled coil protei
321	7	0.5	339	2	B86281	protein F10B6.6 [i	394	7	0.5	426	2	I48379	gene hb protein -

103	7	0.5	116	2	PC4228	N-acetylglucosamin	176	7	0.5	210	2	A34719	nonhistone chromos
104	7	0.5	118	2	T05337	hypothetical prote	177	7	0.5	210	2	S54774	high mobility grou
105	7	0.5	119	2	A03314	homeotic protein m	178	7	0.5	211	2	I53020	G-0/G-1 switch reg
106	7	0.5	121	2	T08717	hypothetical prote	179	7	0.5	211	2	S69929	outer surface prot
107	7	0.5	127	2	A70200	ribosome-binding f	180	7	0.5	213	2	J00075	neuromodulin - gol
108	7	0.5	129	2	H84749	probable squamosa-	181	7	0.5	213	2	J03931	DNA binding protei
109	7	0.5	130	2	G70973	hypothetical prote	182	7	0.5	213	2	G86261	hypothetical prote
110	7	0.5	131	2	PQ0059	T-cell receptor be	183	7	0.5	214	2	G69380	hypothetical prote
111	7	0.5	131	2	E70920	probable moab prot	184	7	0.5	214	2	T51657	myb-related trans
112	7	0.5	131	2	T52597	squamosa promoter	185	7	0.5	214	2	T39559	probable ubiquinol
113	7	0.5	133	2	H69489	SSU ribosomal prot	186	7	0.5	215	1	NSRTH1	nonhistone chromos
114	7	0.5	134	2	E57233	complexin II - hum	187	7	0.5	215	2	I48688	non-histone chromo
115	7	0.5	134	2	D57233	complexin II - mou	188	7	0.5	215	2	T05158	hypothetical prote
116	7	0.5	134	2	JC4226	synaphin - bovine	189	7	0.5	215	2	S29595	calmodulin-related
117	7	0.5	134	2	C57233	complexin II - rat	190	7	0.5	216	2	D48441	calmodulin-like pr
118	7	0.5	134	2	T75416	hypothetical prote	191	7	0.5	217	2	T48642	hypothetical prote
119	7	0.5	136	2	T30045	hypothetical prote	192	7	0.5	219	2	C96510	hypothetical prote
120	7	0.5	138	2	T16473	hypothetical prote	193	7	0.5	220	2	S01063	Hox 2 protein type
121	7	0.5	140	2	G84608	En/spm-like transp	194	7	0.5	221	2	A86163	protein F15K9.22 l
122	7	0.5	141	2	T46015	hypothetical prote	195	7	0.5	223	2	T46043	hypothetical prote
123	7	0.5	142	2	S63373	probable membrane	196	7	0.5	224	2	S51093	transcription acti
124	7	0.5	143	2	T29740	hypothetical prote	197	7	0.5	224	2	JH0456	tumor rejection an
125	7	0.5	144	2	B72280	hypothetical prote	198	7	0.5	227	1	S68150	eosinophil major b
126	7	0.5	149	2	F83964	hypothetical prote	199	7	0.5	228	2	C89347	hypothetical prote
127	7	0.5	152	2	S23311	HMG-Y-related prot	200	7	0.5	229	2	A28329	homeotic protein H
128	7	0.5	154	2	T28434	transcription regu	201	7	0.5	230	2	T39119	protein-L-isoaspar
129	7	0.5	155	2	T05169	Lsd1 protein homol	202	7	0.5	232	2	A69501	hypothetical prote
130	7	0.5	157	2	F71906	hypothetical prote	203	7	0.5	233	2	H82995	hypothetical prote
131	7	0.5	159	2	T44144	B1 protein (import	204	7	0.5	234	1	S15102	conserved hypotet
132	7	0.5	161	2	G89173	hypothetical prote	205	7	0.5	234	2	B44459	eosinophil major b
133	7	0.5	170	1	WNADE5	early E1B 20K prot	206	7	0.5	235	2	F81322	tropomyosin T, fast s
134	7	0.5	170	2	A27853	nonhistone chromos	207	7	0.5	236	2	S33662	hypothetical prote
135	7	0.5	170	2	S69465	hypothetical prote	208	7	0.5	238	1	OQBE74	rpry protein - Bac
136	7	0.5	172	2	D72368	hypothetical prote	209	7	0.5	241	2	A84517	US4 protein - huma
137	7	0.5	173	2	JC6004	superoxide dismuta	210	7	0.5	241	2	T00988	Mutator-like trans
138	7	0.5	173	2	G85771	superoxide dismuta	211	7	0.5	242	2	S19267	hypothetical prote
139	7	0.5	173	2	S73388	hypothetical prote	212	7	0.5	246	2	S38344	anthranilate phosp
140	7	0.5	173	2	T08011	2S seed storage pr	213	7	0.5	247	2	T05226	CDEI-binding prote
141	7	0.5	176	2	S23310	HMG-Y-related prot	214	7	0.5	248	2	T02554	hypothetical prote
142	7	0.5	177	2	T49816	hypothetical prote	215	7	0.5	248	2	T05080	hypothetical prote
143	7	0.5	177	2	C72765	hypothetical prote	216	7	0.5	249	2	T44220	hypothetical prote
144	7	0.5	178	2	T09585	high mobility grou	217	7	0.5	249	2	A84751	probable SF16 prot
145	7	0.5	178	2	T09584	high mobility grou	218	7	0.5	250	2	C71677	hypothetical prote
146	7	0.5	178	2	T51159	HMG protein (import	219	7	0.5	250	2	T26010	hypothetical prote
147	7	0.5	178	2	C83705	hypothetical prote	220	7	0.5	254	2	T01109	hypothetical prote
148	7	0.5	181	1	ERAD23	early E3 20.3K gly	221	7	0.5	256	2	A70637	hypothetical prote
149	7	0.5	181	1	B44057	early E3 20.3K gly	222	7	0.5	256	2	S37926	hypothetical prote
150	7	0.5	181	1	F44057	early E3 20.3K gly	223	7	0.5	259	2	G83675	hypothetical prote
151	7	0.5	181	2	T15117	hypothetical prote	224	7	0.5	259	2	I50727	RNA-binding protei
152	7	0.5	183	2	A84122	pyrazinamidase/nic	225	7	0.5	260	2	T25017	sensory organ home
153	7	0.5	184	2	A34506	23K ubiquitin carr	226	7	0.5	261	2	S10321	hypothetical prote
154	7	0.5	184	2	D84746	hypothetical prote	227	7	0.5	261	2	B86358	31K protein - frog
155	7	0.5	186	2	S30221	nonhistone chromos	228	7	0.5	262	2	S42540	hypothetical prote
156	7	0.5	190	2	A26630	nonhistone chromos	229	7	0.5	262	2	F83498	ribosomal protein
157	7	0.5	191	2	G71710	nucleoplasmin A -	230	7	0.5	262	2	T41536	hypothetical prote
158	7	0.5	193	2	S70260	sco2 protein precu	231	7	0.5	264	2	B96680	yeast anti-silenci
159	7	0.5	193	2	S70266	outer surface prot	232	7	0.5	265	2	H83857	hypothetical prote
160	7	0.5	194	2	H83060	peptidyl-trRNA hydr	233	7	0.5	265	2	JC4970	tryptophan synthas
161	7	0.5	196	2	D82143	conserved hypotet	234	7	0.5	265	2	T46013	slow muscle tropon
162	7	0.5	197	2	S65050	low molecular weig	235	7	0.5	265	2	A43738	hypothetical prote
163	7	0.5	200	2	A26169	nucleoplasmin - Af	236	7	0.5	266	2	H86407	Endo16 protein - s
164	7	0.5	202	2	G71405	hypothetical prote	237	7	0.5	267	1	JQ1323	hypothetical prote
165	7	0.5	202	2	A26301	dopamine- and camp	238	7	0.5	267	2	H86320	DNA excision repai
166	7	0.5	203	2	T50421	nonhistone chromos	239	7	0.5	267	2	H84857	probable MYB47 tra
167	7	0.5	205	2	S26062	nonhistone chromos	240	7	0.5	267	2	E83232	hypothetical prote
168	7	0.5	205	2	H69889	hypothetical prote	241	7	0.5	269	2	T16487	probable ATP-bindi
169	7	0.5	206	2	S30108	S-6b protein - Agr	242	7	0.5	271	2	G85035	hypothetical prote
170	7	0.5	206	2	A69712	mutants block spor	243	7	0.5	271	2	D82314	hypothetical prote
171	7	0.5	208	2	S46510	hypothetical prote	244	7	0.5	273	1	I38886	transcription regu
172	7	0.5	208	2	T05871	hypothetical prote	245	7	0.5	273	2	JG0190	DNA excision repai
173	7	0.5	209	1	NSHUH2	nonhistone chromos	246	7	0.5	276	2	I53161	alpha 2-adrenergic
174	7	0.5	209	2	C69783	NADH dehydrogenase	247	7	0.5	278	2	S48776	hypothetical prote
175	7	0.5	209	2	S89275	CGMP-gated cation	248	7	0.5	279	2	S30766	ASF1 protein - yea

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:04:14 ; Search time 26.41 Seconds
(without alignments)
4012.071 Million cell updates/sec

Title: US-09-512-581-2

Perfect score: 1391

Sequence: 1 MAHSKTRNDGKITYPEVK.....OKGRGRPSKTPSPQPKKNV 1391

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR_68:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	19	1.4	851	2	T00374	hypothetical prote
2	9	0.6	439	2	A57573	telomeric repeat b
3	9	0.6	554	2	T45840	GTPase activating-
4	9	0.6	837	2	T49562	neurofilament-H re
5	9	0.6	1938	1	S06005	myosin alpha heavy
6	9	0.6	1938	2	I49464	alpha cardiac myos
7	8	0.6	81	2	T47289	hypothetical prote
8	8	0.6	87	2	S27275	GTP-binding regula
9	8	0.6	91	1	S00219	ubiquinol--cytochr
10	8	0.6	99	2	A55819	nonhistone chromos
11	8	0.6	147	1	B46315	E4 protein - huma
12	8	0.6	204	2	T05575	hypothetical prote
13	8	0.6	207	2	JC1129	nonhistone chromos
14	8	0.6	231	2	JC1114	high-mobility grou
15	8	0.6	231	2	T49830	hypothetical prote
16	8	0.6	233	1	S13625	eosinophil major b
17	8	0.6	233	2	A23729	somatolactin precu
18	8	0.6	242	1	RHHUT	thyroliberin precu
19	8	0.6	294	2	E84706	hypothetical prote
20	8	0.6	313	2	F84923	pale cress protein
21	8	0.6	316	2	S58719	probable membrane
22	8	0.6	316	2	JC6146	Card protein - Myx
23	8	0.6	325	2	T18283	hypothetical prote
24	8	0.6	339	2	B69436	LSU ribosomal prot
25	8	0.6	341	2	G72775	hypothetical prote
26	8	0.6	351	2	B71604	rRNA methylase (Sp
27	8	0.6	398	2	S06324	dnab protein homol
28	8	0.6	418	2	T15827	hypothetical prote
29	8	0.6	437	2	H86683	prophage pil prote

iron transport sys	1	E70046	442	0.6	8	30
dnab protein homol	2	A37386	451	0.6	8	31
dnab protein homol	2	S01921	451	0.6	8	32
chromogranin A pre	1	A60746	462	0.6	8	33
chromogranin A pre	1	A39868	463	0.6	8	34
probable replicati	2	G71231	468	0.6	8	35
hypothetical prote	2	F84558	507	0.6	8	36
nitrogenase (EC 1.	2	T10830	513	0.6	8	37
hypothetical prote	2	T00828	513	0.6	8	38
tyrosine phosphopr	2	A56110	533	0.6	8	39
involucrin L - dou	1	I36911	544	0.6	8	40
low molecular weig	2	B44841	544	0.6	8	41
receiver-like prot	2	T04653	552	0.6	8	42
related to merzoi	2	T49788	562	0.6	8	43
legumin B - fava b	2	S37241	564	0.6	8	44
general secretion	2	A72329	566	0.6	8	45
prolyl tRNA synthe	2	H86552	568	0.6	8	46
proline--tRNA liga	2	F72070	568	0.6	8	47
unknown protein F2	2	A96766	575	0.6	8	48
Lamin B1 - human	1	VEHULB	586	0.6	8	49
Lamin B - mouse	2	S07720	587	0.6	8	50
hypothetical prote	2	T02398	646	0.6	8	51
amyloid precursor-	2	A46362	653	0.6	8	52
GGDEF family prote	2	F82138	655	0.6	8	53
probable membrane	2	S61693	671	0.6	8	54
hypothetical prote	2	T26819	720	0.6	8	55
Ac-like transposas	2	C84523	730	0.6	8	56
hypothetical prote	2	E96558	765	0.6	8	57
potassium channel	2	S68699	787	0.6	8	58
hypothetical prote	2	B71212	841	0.6	8	59
inositol-polyphosp	2	B57487	938	0.6	8	60
FUN12 protein - ye	2	S70292	1002	0.6	8	61
hypothetical prote	2	T49425	1047	0.6	8	62
hypothetical prote	2	T46444	1062	0.6	8	63
sucrose-phosphate	2	T09837	1081	0.6	8	64
chromatin structur	2	T42204	1082	0.6	8	65
translation initia	2	T43483	1132	0.6	8	66
adenyl cyclase t	2	A6180	1165	0.6	8	67
C-terminal domain-	2	A49201	1166	0.6	8	68
adenylate cyclase	2	T31421	1173	0.6	8	69
adenylate cyclase	2	A47202	1180	0.6	8	70
fibulin-2 precurs	2	A55184	1184	0.6	8	71
hypothetical prote	2	T24415	1231	0.6	8	72
X-linked nuclear p	2	I54367	1298	0.6	8	73
hypothetical prote	2	T26517	1634	0.6	8	74
helicase II - huma	2	I38614	1841	0.6	8	75
hypothetical prote	2	T00391	1736	0.6	8	76
probable heat shoc	2	D96796	1871	0.6	8	77
myosin heavy chain	2	I48153	1934	0.6	8	78
myosin beta heavy	1	S06006	1935	0.6	8	79
adenylate cyclase	1	D42088	2248	0.6	8	80
ryanodine-binding	2	B54161	4868	0.6	8	81
hypothetical prote	2	T00418	35	0.5	7	82
hypothetical prote	2	E70210	41	0.5	7	83
Ku antigen, 70K -	2	S25149	46	0.5	7	84
hypothetical prote	2	T06013	75	0.5	7	85
ubiquinol--cytochr	2	CC8011	78	0.5	7	86
hypothetical prote	2	T48398	81	0.5	7	87
hypothetical prote	2	T45132	91	0.5	7	88
hypothetical prote	2	T46086	95	0.5	7	89
ribosomal protein	2	H69091	101	0.5	7	90
parathymosin - rat	2	B31512	102	0.5	7	91
homeotic protein R	2	A27471	105	0.5	7	92
zinc-binding prote	2	S20422	106	0.5	7	93
LSU ribosomal prot	2	C69436	106	0.5	7	94
protein F1504.29 [2	C86477	107	0.5	7	95
prothymosin alpha	1	TNBOA1	109	0.5	7	96
prothymosin alpha	2	F72666	109	0.5	7	97
prothymosin alpha	1	TNBUA	110	0.5	7	98
prothymosin alpha	2	S15073	111	0.5	7	99
hypothetical prote	1	TNRTA	112	0.5	7	100
DNA binding protei	2	T18436	112	0.5	7	101
	2	S57448	112	0.5	7	102

ACCESSION	AL358892
VERSION	AL358892.12
KEYWORDS	GI:11493289
SOURCE	HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULFILLTOP.
ORGANISM	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
	1 (bases 1 to 195380)
	Wall, M.
REFERENCE	Direct Submission
AUTHORS	Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL	CG10 ISA. UK. E-mail enquiries: humquery@sanger.ac.uk Clone
	requests: clonerequest@sanger.ac.uk
COMMENT	On Nov 30, 2000 this sequence version replaced gi:11414592.

```

Center : UK Medical Research Council
Center code : UK-MRC
Web site: http://mroseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: dm41766
----- Summary Statistics
Sequencing program: XMAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 190612 bases at least Q40
Consensus quality: 192224 bases at least Q30
Consensus quality: 193166 bases at least Q20
Insert size: 194680; sum-of-contigs
Insert size: 191947; 3.7% error; agarose-fp
Quality coverage: 7.33x in Q20 bases; sum-Of-contigs quality
coverage: 8.85x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

FEATURES	
source	location/offsets
	1. 195380
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/chromosome="5"
	/clone="RP21-417G6"
	/clone_1fb="RPC1-21"
misc_feature	1. 2333
/note="assembly_fragment:02826	
fragment_Chain:1	
2433. .4472	
/note="assembly_fragment:00159	
fragment_Chain:1	
4573. 6623	
/note="assembly_fragment:00292"	
6724. .8757	
/note="assembly_fragment:02097"	
8858. .11498	
/note="assembly_fragment:02560"	
11599. .13673	
/note="assembly_fragment:03440"	
13774. .16148	
/note="assembly_fragment:03977"	
16249. .195380	
/note="assembly_fragment:04183"	
BASE COUNT	55503 a 40458 c 40664 g 58048 t 707 others
ORIGIN	

Query Match	4.38;	Score 179.6;	DB 80;	Length 195380;
Best Local Similarity	91.0%;	Pred. No. 8.3e-27;		
Matches 191; Conservative	0;	Mismatches 19;	Indels 0;	Gaps 0;

Qy 104 taagatggttctgaacaactttatgatatygaccagaccttgaaagaanaaggagc 163
+
Db 168033 TATGATGGCTGTAAACCTTCATGCAGCATGGAACCAAGACTCGAAGAAGAAAAGAAC 168092

QY 164 ttctttaaaccctgagctttacaaccttgctcaagtttttttccaagacatccctggtaaag 223
|||||
Db 168093 TTTATCTTAACCTGACCTTTACATCTTGCTCTTGACTCTTCTTCTCAACGATCTGATTAAG 168152
|||||
QY 224 atgttcgcttactgtagccctgcgcctctgctgataatttccaagatttaatgcctcgaag 283
|||||
Db 168153 ATGTTTCGTTTACGTGTGGCTTGCTGCGCTTGCTGATATTTCAGAGATTTAATGCTCTCAGG 168212
|||||
QY 284 ctcccttaacacatccctctgataaactaaag 313
|||||
Db 168213 CTCCTTACACGCTTCGCGATTAACCTTAAGG 168242
|||||

```
Search completed: September 25, 2001, 20:35:12
Job time: 22075 sec
```


| QY | 104 | taaagatggtgtgnaaaccttltatgatalggaaccgagacttcgaaagaagaagagac | 163 |
|------------|--|---|-------------|
| DB | 16393 | TATGATGATGTTGTGAAACCTTTATGATATGAGACCGAGACTCTGAAGAAACAAAAGAGC | 16452 |
| QY | 164 | ttatataaacctagcctttacatccttgcttcagatlttttctcaacgacatcctgtgtaaag | 223 |
| DB | 16453 | TTTATTAAACCTGAGCTTTACATCTGTGGTTCAGATTTTTCCTCAAGCATCTGATTAAG | 16512 |
| QY | 224 | atgttcgcttactcgtgtagcctgcgtgccttgcgtatatttcaagatltatgcttcgaaag | 283 |
| DB | 16513 | ATGTTCGCTTACTCGTAGCCCTGCCTCTCTGATATTTTCAGGATTTATGCTCTCTGAAG | 16572 |
| QY | 284 | ctcctcacacatccctcgtataactaaagg | 313 |
| DB | 16573 | CTCCTTACACATCCCTGTATTAACATAAGG | 16602 |
| RESULT | 14 | | |
| LOCUS | AL353724 | 163642 bp | DNA |
| DEFINITION | | | HNG |
| | | | 23-JAN-2001 |
| ACCESSION | AL353724 | | |
| VERSION | AL353724.3 | GI:9930876 | |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| REFERENCE | 1 (bases 1 to 163642) | | |
| AUTHORS | Burton,J. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Genome Center | | |
| COMMENT | | | |
| | Center: Sanger Centre | | |
| | Center code: SC | | |
| | Web site: http://www.sanger.ac.uk | | |
| | Contact: humquerry@sanger.ac.uk | | |
| | Project Information | | |
| | Center project name: B446113 | | |
| | Summary Statistics | | |
| | Assembly program: XGAP4; version 4.5 | | |
| | Sequencing vector: plasmid; L08752; 100% of reads | | |
| | Chemistry: Dye-terminator Big Dye; 100% of reads | | |
| | Consensus quality: 159306 bases at least Q40 | | |
| | Consensus quality: 161143 bases at least Q30 | | |
| | Consensus quality: 162084 bases at least Q20 | | |
| | Insert size: 162842; sum-of-contigs | | |
| | Insert size: 163676; 2.5% error; agarose-fp | | |
| | Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality | | |
| | coverage: 4.42x in Q20 bases; agarose-fp | | |
| | NOTE: This is a 'working draft' sequence. It currently | | |
| | * consists of 9 contigs. The true order of the pieces | | |
| | * is not known and their order in this sequence record is | | |
| | * arbitrary. Gaps between the contigs are represented as | | |
| | * runs of N, but the exact sizes of the gaps are unknown. | | |
| | * This record will be updated with the finished sequence | | |
| | * as soon as it is available and the accession number will | | |
| | * be preserved. | | |
| | * 1 20261: contig of 20261 bp in length | | |
| | * 20262 20361: gap of 100 bp | | |
| | * 20362 38140: contig of 17779 bp in length | | |
| | * 38141 38240: gap of 100 bp | | |
| | * 38241 85018: contig of 46778 bp in length | | |
| | * 85019 85118: gap of 100 bp | | |
| | * 85119 93385: contig of 8267 bp in length | | |
| | * 93386 93485: gap of 100 bp | | |
| | * 93486 96470: contig of 2985 bp in length | | |

```

* 96471 96570: gap of 100 bp
* 96571 137301: contig of 40731 bp in length
* 137302 137401: gap of 100 bp
* 137402 147035: contig of 9634 bp in length
* 147036 147135: gap of 100 bp
* 147136 149710: contig of 2575 bp in length
* 149711 149810: gap of 100 bp
* 149811 163642: contig of 13832 bp in length.
FEATURES
    source
        1. 163642
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="13"
            /clone="RP11-448113"
            /clone_1fb="RPC1-11.2"
            1. 20261
                /note="assembly_fragment:00330
                clone_end:77
                vector_side:left"
                20362..38140
                    /note="assembly_fragment:00335
                    fragment_chain:1"
                    38241..85018
                        /note="assembly_fragment:01397
                        fragment_chain:1"
                        85119..93385
                            /note="assembly_fragment:01083
                            fragment_chain:1"
                            93486..96470
                                /note="assembly_fragment:00197
                                fragment_chain:1"
                                96571..137301
                                    /note="assembly_fragment:00284
                                    fragment_chain:2"
                                    137402..147035
                                        /note="assembly_fragment:00874
                                        fragment_chain:2"
                                        147136..149710
                                            /note="assembly_fragment:00391
                                            fragment_chain:3"
                                            149811..163642
                                                /note="assembly_fragment:00398
                                                fragment_chain:3"
BASE COUNT  54756 a 31368 c 29374 g 47337 t      807 others
ORIGIN
Query Match          5.0%: Score 206.8; DB 79; Length 163642;
Best Local Similarity 99.0%: Pred. No.2.1e-32;
Matches 208; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 104  taaagaatgcttgtaaaactttatgatgatgataagaccagagactctgaagaagaaaaagagac 163
      |||
Db 32941 TATGATGAGTGTGAAAACCTTTATGATGATGAGACCAAGACTGTGAAGAAAAGAGAGC 32882
QY 164  ttatttaacctgaatctttaaactcttgcttcaagattttttctcaagaatcctgtaag 223
      |||
Db 32881 TTTATTTTAAACCTACCTTTACATCTTTCGCTTCAGATTTTTTCTCAAGCATCCTGATTAAG 32822
QY 224  atgctgcgttactgtaagcctctgcctctgctgcatatttcaagaatctatgctcctgaag 283
      |||
Db 32821 ANGTCGCTACTACTGTGACCTGCTGCTGCTGCGATTTTTCAGATTTATATGCTCTCGTAAG 32762
QY 284  ctctctaacatccctctataaactaaag 313
      |||
Db 32761 CTCTTACACATCCCTGATTAACCTAAAG 32732
RESUT 15
AL358892 195380 bp DNA HTG 06-MAR-2001
LOCUS Mus musculus chromosome 5 clone RP21-417G6, *** SEQUENCING IN
DEFINITION PROGRAMS ***. In unordered pieces.

```



```
repeat_region 9724..9876
/partial
/note="Alu repeat: matches 158..1 of consensus"
repeat_region 10567..10594
/note="14 copies of 2 mer 93 & conserved"
repeat_region 10616..10900
/partial
/note="Alu repeat: matches 300..1 of consensus"
repeat_region 10960..10985
/note="13 copies of 2 mer 92 & conserved"
repeat_region 11912..12100
/partial
/note="Alu repeat: matches 268..65 of consensus"
repeat_region 17050..17359
/note="Alu repeat: matches 1..306 of consensus"
repeat_region 17639..17888
/partial
/note="Alu repeat: matches 34..306 of consensus"
repeat_region 19151..19980
/note="11 element fragment"
repeat_region 19981..20264
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 20238..21190
/note="11 element fragment"
repeat_region 21199..21488
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 21538..22226
/note="11 element fragment"
repeat_region 25696..25989
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 26902..27167
/note="2 copies of 133 mer 93 & conserved"
repeat_region 27850..27930
/note="11 element fragment"
repeat_region 28002..28043
/note="11 element fragment"
repeat_region 28333..28780
/note="11 element fragment"
repeat_region 29176..29449
/note="Alu repeat: matches 1..306 of consensus"
repeat_region 29904..30192
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 30354..30663
/note="11 element fragment"
repeat_region 30683..30972
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 30977..31204
/note="11 element fragment"
repeat_region 34819..35121
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 37059..37282
/note="MLR2B2 element fragment"
repeat_region 37453..37506
/note="MLR2B2 element fragment"
repeat_region 38414..38706
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 39516..39563
/note="24 copies of 2 mer 81 & conserved"
repeat_region 41244..41535
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 43994..44242
/partial
/note="Alu repeat: matches 44..308 of consensus"
repeat_region 44369..44677
/note="match: 217151 DNA segment containing (CA) repeat"
/note="16 copies of CA 100% conserved; Polymorphic by
comparison with 217151"

repeat_region 44702..44889
/note="2 copies of 94 mer 85 & conserved"
repeat_region 45935..46229
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 46910..46988
/note="11 element fragment"
repeat_region 47081..47119
/note="11 element fragment"
repeat_region 47168..47431
/note="11 element fragment"
repeat_region 47615..47902
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 47922..48213
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 49426..49717
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 50248..50533
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 51097..51380
/note="Alu repeat: matches 1..304 of consensus"
repeat_region 62318..62543
/partial
/note="Alu repeat: matches 302..66 of consensus"
repeat_region 63358..63613
/partial
/note="Alu repeat: matches 306..34 of consensus"
repeat_region 66191..66236
/note="23 copies of 2 mer 80 & conserved"
repeat_region 68771..69062
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 70282..70559
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 70590..70881
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 75593..75883
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 75974..76251
/partial
/note="Alu repeat: matches 294..1 of consensus"
repeat_region 76551..76608
/note="MIR element fragment"
repeat_region 78482..78832
/note="MER11B element fragment"
repeat_region 78574..78743
/note="MER11A element fragment"
repeat_region 78941..79021
/note="MER11B element fragment"
repeat_region 79249..79709
/note="MER11A element fragment"
repeat_region 80233..80302
/note="SVA element fragment"
repeat_region 80355..80535
/partial
/note="Alu repeat: matches 213..20 of consensus"
repeat_region 80563..80619
/partial
/note="Alu repeat: matches 277..221 of consensus"
repeat_region 80651..82294
/note="SVA element fragment"
repeat_region 82753..83043
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 83149..83296

Query Match 5.0%; Score 206.8; DB 92; Length 113704;
Best Local Similarity 99.0%; Pred. No. 2.1e-32;
Matches 208; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```


Oy 2729 ctatcaacgatgtaatgctatccaagtaagacaagtgtttgccagaacttcacaaagcc 2788
||| | |||||||
Db 4412 cttttacagcatgaatgcctatcacaagtaagacaaagtgtttgccagaaccttcacaaaggcc 4477

```

/note="Alu repeat: matches 305. .210 of consensus"

```


Db 481 GAAGGAACCTTATGTAGACACTGCACTGAGACTGGAACCAATATTATTAATT 540
 QY 3485 caaatcaactctctcctggaagataaaggagagctgtagtctgca 3533
 Db 541 CAGAGCTGAACCTTCACCGAATGCATCAAGGAACAGAGTTGAGA 589
 RESULT 12
 AC068224
 LOCUS Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
 DEFINITION SEQUENCE SAMPLING.
 AC068224
 VERSION AC068224.1 GI:7671284
 KEYWORDS HTG: HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 54398)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 3, clone RP11-660H19
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 54398)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
 Campione,A., Castle,A., Choquel,X., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardina,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-pierre,N., Grant,G., Hagos,B., Harford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lebecky,J.,
 Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
 Melidim,J., Menuez,L., Mihova,T., Miranda,C., Mlenge,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Teste,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 DIRECT SUBMISSION
 Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10161
 Center clone name: 660_H19

 * NOTE: This record contains 68 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 1 685: contig of 685 bp in length

686 785: gap of 100 bp
 786 1513: contig of 728 bp in length
 1514 1613: gap of 100 bp
 1614 2354: contig of 741 bp in length
 2355 2454: gap of 100 bp
 2455 3184: contig of 730 bp in length
 3185 3284: gap of 100 bp
 3285 3959: contig of 675 bp in length
 3960 4059: gap of 100 bp
 4060 4746: contig of 687 bp in length
 4747 4846: gap of 100 bp
 4847 5564: contig of 718 bp in length
 5565 5664: gap of 100 bp
 5665 6360: contig of 696 bp in length
 6361 6460: gap of 100 bp
 6461 7132: contig of 672 bp in length
 7133 7232: gap of 100 bp
 7233 7937: contig of 705 bp in length
 7938 8037: gap of 100 bp
 8038 8749: contig of 712 bp in length
 8750 8849: gap of 100 bp
 8850 9522: contig of 673 bp in length
 9523 9622: gap of 100 bp
 9623 10356: contig of 734 bp in length
 10357 10456: gap of 100 bp
 10457 11181: contig of 725 bp in length
 11182 11281: gap of 100 bp
 11282 11975: contig of 694 bp in length
 11976 12075: gap of 100 bp
 12076 12785: contig of 710 bp in length
 12786 12885: gap of 100 bp
 12886 13557: contig of 672 bp in length
 13558 13657: gap of 100 bp
 13658 14347: contig of 690 bp in length
 14348 14447: gap of 100 bp
 14448 15132: contig of 685 bp in length
 15133 15232: gap of 100 bp
 15233 15912: contig of 680 bp in length
 15913 16012: gap of 100 bp
 16013 16731: contig of 719 bp in length
 16732 16831: gap of 100 bp
 16832 17506: contig of 675 bp in length
 17507 17606: gap of 100 bp
 17607 18339: contig of 733 bp in length
 18340 18439: gap of 100 bp
 18440 19109: contig of 670 bp in length
 19110 19209: gap of 100 bp
 19210 19900: contig of 691 bp in length
 19901 20000: gap of 100 bp
 20001 20693: contig of 693 bp in length
 20694 20793: gap of 100 bp
 20794 21464: contig of 671 bp in length
 21465 21564: gap of 100 bp
 21565 22245: contig of 681 bp in length
 22246 22345: gap of 100 bp
 22346 23049: contig of 704 bp in length
 23050 23149: gap of 100 bp
 23150 23857: contig of 708 bp in length
 23858 23957: gap of 100 bp
 23958 24685: contig of 728 bp in length
 24686 24785: gap of 100 bp
 24786 25499: contig of 714 bp in length
 25500 25599: gap of 100 bp
 25600 26302: contig of 703 bp in length
 26303 26402: gap of 100 bp
 26403 27108: contig of 706 bp in length
 27109 27208: gap of 100 bp
 27209 27910: contig of 702 bp in length
 27911 28010: gap of 100 bp
 28011 28720: contig of 710 bp in length
 28721 28820: gap of 100 bp
 28821 29484: contig of 664 bp in length
 29485 29584: gap of 100 bp

| | | | | | |
|---|-------|--------|------------------|-----------|--|
| * | 49510 | 49609: | gap of | 100 bp | |
| * | 49610 | 50331: | contig of 722 bp | in length | |
| * | 50332 | 50431: | gap of 100 bp | | |
| * | 50432 | 51135: | contig of 704 bp | in length | |
| * | 51136 | 51235: | gap of 100 bp | | |
| * | 51236 | 51981: | contig of 746 bp | in length | |
| * | 51982 | 52081: | gap of 100 bp | | |
| * | 52082 | 52783: | contig of 702 bp | in length | |
| * | 52784 | 52883: | gap of 100 bp | | |
| * | 52884 | 53580: | contig of 697 bp | in length | |
| * | 53581 | 53680: | gap of 100 bp | | |
| * | 53681 | 54398: | contig of 718 bp | in length | |

/organism="Homo sapiens"

| | | | | |
|---------------------------|--------|--------------------|------------|---------------|
| Query Match | 9.38; | Score 389.8; | DB 73; | Length 54398; |
| Best Local Similarity | 83.78; | Pred. No. 4.9e-70; | | |
| Matches 494; Conservative | 0; | Mismatches 3; | Indels 93; | Gaps 1 |

| | | | |
|----|-------|--|-------|
| Oy | 3677 | agagaaatcagatcgaacttgacaaattgtaacagaaacccaataagca | 3736 |
| Db | 29484 | agagaaattagctatgcgatcgaacttaagtttgctacagagacgaacttaaggca | 29425 |

[illegible]

3797 aggaagaagagcctcaagaagatatattgaanaatcgaagtgaacagaatgtagcgccaa 3856

| | | | |
|----|-------|---|-------|
| Ov | 3857 | aaagatggttaaaaagggggaaccccccgcctcgttttgttcaaacacccccccccccccc | 2016 |
| Db | 29364 | AGGAAAGAAGGCCTCAAGATGATATTATTAGAAAATGAAGATGACAGAAATAGTCGGCCAA | 29305 |

Db 29304 AAAAGGCTAAAGAGGGCCGACACCACCAAACTCTTGGTGAAGTACACCAAAAGAGAGC 29245

DB 29344 CACCAATGAAACTTCTTAAAAAGGAGCAAAAAAATCTGGACCTCCAGCAACAGAG 29185
DY 3917 caacaatgaaactcttaaaaaaagaagcaaaaaaaatctgacatccagcaccagag 3976

OY 3977 agggaggaagaagaagaacaaagtgyaatatacygaacagaagtccaaaagcaaacgc 4036
|||||||

| | | | | | |
|------|-----------|--------|----------|-------|------|
| 4037 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4062 |
| 4038 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4063 |
| 4039 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4064 |
| 4040 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4065 |
| 4041 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4066 |
| 4042 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4067 |
| 4043 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4068 |
| 4044 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4069 |
| 4045 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4070 |
| 4046 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4071 |
| 4047 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4072 |
| 4048 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4073 |
| 4049 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4074 |
| 4050 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4075 |
| 4051 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4076 |
| 4052 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4077 |
| 4053 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4078 |
| 4054 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4079 |
| 4055 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4080 |
| 4056 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4081 |
| 4057 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4082 |
| 4058 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4083 |
| 4059 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4084 |
| 4060 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4085 |
| 4061 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4086 |
| 4062 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4087 |
| 4063 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4088 |
| 4064 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4089 |
| 4065 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4090 |
| 4066 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4091 |
| 4067 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4092 |
| 4068 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4093 |
| 4069 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4094 |
| 4070 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4095 |
| 4071 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4096 |
| 4072 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4097 |
| 4073 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4098 |
| 4074 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4099 |
| 4075 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4100 |
| 4076 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4101 |
| 4077 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4102 |
| 4078 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4103 |
| 4079 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4104 |
| 4080 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4105 |
| 4081 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4106 |
| 4082 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4107 |
| 4083 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4108 |
| 4084 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4109 |
| 4085 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4110 |
| 4086 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4111 |
| 4087 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4112 |
| 4088 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4113 |
| 4089 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4114 |
| 4090 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4115 |
| 4091 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4116 |
| 4092 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4117 |
| 4093 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4118 |
| 4094 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4119 |
| 4095 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4120 |
| 4096 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4121 |
| 4097 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4122 |
| 4098 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4123 |
| 4099 | accgaqtlc | aaqaaq | cacacaaq | ----- | 4124 |

Db 29124 ACCGAGTGTCAAGAGAGACAGCAGAGGTAAGCATGTGTAAGCTCTAACTGCATCTGTT 29065

| | | | |
|----|-------|---|-------|
| Db | 29064 | TCGTTACTATATTATAAANCATTAATTGATGCTATCCACATTTGGGCTTCCCCCAAGCA | 29005 |
| | 4063 | -----a | 4063 |

4064 gagcgaatctctgaatctagtgcaattgaatccacacagtcacaccacagaagaagac 4123

4124 gaggaagaccatcaaaaacgccatcaccatcacacccaataaaatgtg 4173

Db 28944 GAGAGAGACCATCAAAAAAGCCCATCACCATCACAACCAAAAAAATGTG 28895

```

RESULT 11
AK021757
-----

```

DEFINITION Homo sapiens cDNA FLJ11635 fis, clone HEMBA1005019, highly similar to Homo sapiens mRNA for KIAA0648 protein.

```

VERSION      AK021757.1  GI:10433003
KEYWORDS     oligo capping; fls (full insert sequence).
COMMENT      Homo sapiens embryonic 10 weeks fetal testes
              cDNA; full-length cDNA.

```

ORGANISM
mRNA, clone_lib:HEMB1 clone:HEMB1005019.
Homo sapiens
Phylogeny: Metazoa: Chordata: Vertebrata: Euteleostomi:
Eulavota: Metazoa: Chordata: Vertebrata: Euteleostomi:

REFERENCE

AUTHORS

REFERENCE
AUTHORS
TITLE

Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
1 (sales)
I (sales), Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Iisoga, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Haru, H.,
Nishikawa, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takuchi, K.,
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuko, Y. and Oshima, A.
NEDO human cDNA sequencing project

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

JOURNAL
Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao
Isogai, Heliix Research Institute, Genomics Laboratory, 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomicsehr1.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human CDNA sequencing project supported by Ministry of

NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .2212 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1005019"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/issue_type="whole embryo, mainly head"
/notes="Cloning vector: pME18SFL3"
BASE COUNT      751 a      385 c      430 g      646 t
ORIGIN

```

| | | | | |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match | 5.6%; | Score 232.2; | DB 89; | Length 2212; |
| Best Local Similarity | 62.1%; | Pred. No. 1.2e-37; | | |
| Matches 366; | Conservative 0; | Mismatches 223; | Indels 0; | Gaps 0; |

Oy 2945 aatattgccccttcacacagatgattgttccatatacaatccacttcggacatg 3004
 1 AATTATTACTGTTGCCGTGATATAGTATTCATACATGATTTCACTGCTAGCCATG 60

QY 3005 acccagatla^glcaaa^gta^cacagata^ttgacaac^ttaaa^gat^glttaa^gaat^gtc^ttt 3064

QY 3065 ggattgtctcgaaatatlaatggtcfaaaaatgaanaataacagtcacgcgtttatcaga 3124

Db 121 GGTTCACTGCTTCAAGTTTAAATGACAAAGAAATGAAAACCAATAGCCATGCCCTTATGAACA 180

Db 181 AGATGGCAGAGAACATCAAGTTAAACGAGAGATGCCAGTCTCCAGATGAATCCAGACAA 240

3185 atgnaaacctgacacctgtgtgtgtgtgtgcatagaatcatcatcgtccaagagtacta 3244
 3190 |||||
 3195 |||||
 3200 |||||
 3205 |||||
 3210 |||||
 3215 |||||
 3220 |||||
 3225 |||||
 3230 |||||
 3235 |||||
 3240 |||||
 3245 |||||
 3250 |||||
 3255 |||||
 3260 |||||
 3265 |||||
 3270 |||||
 3275 |||||
 3280 |||||
 3285 |||||
 3290 |||||
 3295 |||||
 3300 |||||
 3305 |||||
 3310 |||||
 3315 |||||
 3320 |||||
 3325 |||||
 3330 |||||
 3335 |||||
 3340 |||||
 3345 |||||
 3350 |||||
 3355 |||||
 3360 |||||
 3365 |||||
 3370 |||||
 3375 |||||
 3380 |||||
 3385 |||||
 3390 |||||
 3395 |||||
 3400 |||||
 3405 |||||
 3410 |||||
 3415 |||||
 3420 |||||
 3425 |||||
 3430 |||||
 3435 |||||
 3440 |||||
 3445 |||||
 3450 |||||
 3455 |||||
 3460 |||||
 3465 |||||
 3470 |||||
 3475 |||||
 3480 |||||
 3485 |||||
 3490 |||||
 3495 |||||
 3500 |||||
 3505 |||||
 3510 |||||
 3515 |||||
 3520 |||||
 3525 |||||
 3530 |||||
 3535 |||||
 3540 |||||
 3545 |||||
 3550 |||||
 3555 |||||
 3560 |||||
 3565 |||||
 3570 |||||
 3575 |||||
 3580 |||||
 3585 |||||
 3590 |||||
 3595 |||||
 3600 |||||
 3605 |||||
 3610 |||||
 3615 |||||
 3620 |||||
 3625 |||||
 3630 |||||
 3635 |||||
 3640 |||||
 3645 |||||
 3650 |||||
 3655 |||||
 3660 |||||
 3665 |||||
 3670 |||||
 3675 |||||
 3680 |||||
 3685 |||||
 3690 |||||
 3695 |||||
 3700 |||||
 3705 |||||
 3710 |||||
 3715 |||||
 3720 |||||
 3725 |||||
 3730 |||||
 3735 |||||
 3740 |||||
 3745 |||||
 3750 |||||
 3755 |||||
 3760 |||||
 3765 |||||
 3770 |||||
 3775 |||||
 3780 |||||
 3785 |||||
 3790 |||||
 3795 |||||
 3800 |||||
 3805 |||||
 3810 |||||
 3815 |||||
 3820 |||||
 3825 |||||
 3830 |||||
 3835 |||||
 3840 |||||
 3845 |||||
 3850 |||||
 3855 |||||
 3860 |||||
 3865 |||||
 3870 |||||
 3875 |||||
 3880 |||||
 3885 |||||
 3890 |||||
 3895 |||||
 3900 |||||
 3905 |||||
 3910 |||||
 3915 |||||
 3920 |||||
 3925 |||||
 3930 |||||
 3935 |||||
 3940 |||||
 3945 |||||
 3950 |||||
 3955 |||||
 3960 |||||
 3965 |||||
 3970 |||||
 3975 |||||
 3980 |||||
 3985 |||||
 3990 |||||
 3995 |||||
 4000 |||||
 4005 |||||
 4010 |||||
 4015 |||||
 4020 |||||
 4025 |||||
 4030 |||||
 4035 |||||
 4040 |||||
 4045 |||||
 4050 |||||
 4055 |||||
 4060 |||||
 4065 |||||
 4070 |||||
 4075 |||||
 4080 |||||
 4085 |||||
 4090 |||||
 4095 |||||
 4100 |||||
 4105 |||||
 4110 |||||
 4115 |||||
 4120 |||||
 4125 |||||
 4130 |||||
 4135 |||||
 4140 |||||
 4145 |||||
 4150 |||||
 4155 |||||
 4160 |||||
 4165 |||||
 4170 |||||
 4175 |||||
 4180 |||||
 4185 |||||
 4190 |||||
 4195 |||||
 4200 |||||
 4205 |||||
 4210 |||||
 4215 |||||
 4220 |||||
 4225 |||||
 4230 |||||
 4235 |||||
 4240 |||||
 4245 |||||
 4250 |||||
 4255 |||||
 4260 |||||
 4265 |||||
 4270 |||||
 4275 |||||
 4280 |||||
 4285 |||||
 4290 |||||
 4295 |||||
 4300 |||||
 4305 |||||
 4310 |||||
 4315 |||||
 4320 |||||
 4325 |||||
 4330 |||||
 4335 |||||
 4340 |||||
 4345 |||||
 4350 |||||
 4355 |||||
 4360 |||||
 4365 |||||
 4370 |||||
 4375 |||||
 4380 |||||
 4385 |||||
 4390 |||||
 4395 |||||
 4400 |||||
 4405 |||||
 4410 |||||
 4415 |||||
 4420 |||||
 4425 |||||
 4430 |||||
 4435 |||||
 4440 |||||
 4445 |||||
 4450 |||||
 4455 |||||
 4460 |||||
 4465 |||||
 4470 |||||
 4475 |||||
 4480 |||||
 4485 |||||
 4490 |||||
 4495 |||||
 4500 |||||
 4505 |||||
 4510 |||||
 4515 |||||
 4520 |||||
 4525 |||||
 4530 |||||
 4535 |||||
 4540 |||||
 4545 |||||
 4550 |||||
 4555 |||||
 4560 |||||
 4565 |||||
 4570 |||||
 4575 |||||
 4580 |||||
 4585 |||||
 4590 |||||
 4595 |||||
 4600 |||||
 4605 |||||
 4610 |||||
 4615 |||||
 4620 |||||
 4625 |||||
 4630 |||||
 4635 |||||
 4640 |||||
 4645 |||||
 4650 |||||
 4655 |||||
 4660 |||||
 4665 |||||
 4670 |||||
 4675 |||||
 4680 |||||
 4685 |||||
 4690 |||||
 4695 |||||
 4700 |||||
 4705 |||||
 4710 |||||
 4715 |||||
 4720 |||||
 4725 |||||
 4730 |||||
 4735 |||||
 4740 |||||
 4745 ||||

3245 catacagtttggaaatctctctaagaaccggtactactaccagctcglttcttcaactcaactg 3304

| | | | |
|----|------|---|------|
| 0y | 3365 | acaaagaatttcagtaacacccaanaatlattctgcctcctcgaatgaaatcatttlltcaactc | 3364 |
| DD | 301 | tgttcacatgcagatttcaccctaaaggaccagttccttcccaatgaaaaattttttacacacacactg | 360 |

Db 361 AAAAGACTTCTGTAACGATTAAGACTTATATTCCAGAAGAGACACAGTAGTACTTCTGTAA 420

[illegible]

QY 3425 gcaagcaatctcagaccaatcatcagaaatgynaactgtaagcaatgcaagcagcagct 3484

COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10161

Center clone name: 660_H_19

* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 685: contig of 685 bp in length
* 686 785: gap of 100 bp
* 786 1513: contig of 728 bp in length
* 1514 1613: gap of 100 bp
* 1614 2354: contig of 741 bp in length
* 2355 2454: gap of 100 bp
* 2455 3184: contig of 730 bp in length
* 3185 3284: gap of 100 bp
* 3285 3959: contig of 675 bp in length
* 3960 4059: gap of 100 bp
* 4060 4746: contig of 687 bp in length
* 4747 4846: gap of 100 bp
* 4847 5564: contig of 718 bp in length
* 5565 5664: gap of 100 bp
* 5665 6360: contig of 696 bp in length
* 6361 6460: gap of 100 bp
* 6461 7132: contig of 672 bp in length
* 7133 7232: gap of 100 bp
* 7233 7937: contig of 705 bp in length
* 7938 8037: gap of 100 bp
* 8038 8749: contig of 712 bp in length
* 8750 8849: gap of 100 bp
* 8850 9522: contig of 673 bp in length
* 9523 9622: gap of 100 bp
* 9623 10356: contig of 734 bp in length
* 10357 10456: gap of 100 bp
* 10457 11181: contig of 725 bp in length
* 11182 11281: gap of 100 bp
* 11282 11975: contig of 694 bp in length
* 11976 12075: gap of 100 bp
* 12076 12785: contig of 710 bp in length
* 12786 12885: gap of 100 bp
* 12886 13557: contig of 672 bp in length
* 13558 13657: gap of 100 bp
* 13658 14347: contig of 690 bp in length
* 14348 14447: gap of 100 bp
* 14448 15132: contig of 685 bp in length
* 15133 15232: gap of 100 bp
* 15233 15912: contig of 680 bp in length
* 15913 16012: gap of 100 bp
* 16013 16731: contig of 719 bp in length
* 16732 16831: gap of 100 bp
* 16832 17506: contig of 675 bp in length
* 17507 17606: gap of 100 bp
* 17607 18339: contig of 733 bp in length
* 18340 18439: gap of 100 bp
* 18440 19109: contig of 670 bp in length
* 19110 19209: gap of 100 bp
* 19210 19900: contig of 691 bp in length
* 19901 20000: gap of 100 bp

* 20001 20693: contig of 693 bp in length
* 20694 20793: gap of 100 bp
* 20794 21464: contig of 671 bp in length
* 21465 21564: gap of 100 bp
* 21565 22245: contig of 681 bp in length
* 22246 22345: gap of 100 bp
* 22346 23049: contig of 704 bp in length
* 23050 23149: gap of 100 bp
* 23150 23857: contig of 708 bp in length
* 23858 23957: gap of 100 bp
* 23958 24685: contig of 728 bp in length
* 24686 24785: gap of 100 bp
* 24786 25499: contig of 714 bp in length
* 25500 25599: gap of 100 bp
* 25600 26302: contig of 703 bp in length
* 26303 26402: gap of 100 bp
* 26403 27108: contig of 706 bp in length
* 27109 27208: gap of 100 bp
* 27209 27910: contig of 702 bp in length
* 27911 28010: gap of 100 bp
* 28011 28720: contig of 710 bp in length
* 28721 28820: gap of 100 bp
* 28821 29484: contig of 664 bp in length
* 29485 29584: gap of 100 bp
* 29585 30328: contig of 744 bp in length
* 30329 30428: gap of 100 bp
* 30429 31111: contig of 683 bp in length
* 31112 31211: gap of 100 bp
* 31212 31931: contig of 720 bp in length
* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length
* 39913 40012: gap of 100 bp
* 40013 40712: contig of 700 bp in length
* 40713 40812: gap of 100 bp
* 40813 41514: contig of 702 bp in length
* 41515 41614: gap of 100 bp
* 41615 42294: contig of 680 bp in length
* 42295 42394: gap of 100 bp
* 42395 43094: contig of 700 bp in length
* 43095 43194: gap of 100 bp
* 43195 43915: contig of 721 bp in length
* 43916 44015: gap of 100 bp
* 44016 44740: contig of 725 bp in length
* 44741 44840: gap of 100 bp
* 44841 45533: contig of 693 bp in length
* 45534 45633: gap of 100 bp
* 45634 46330: contig of 697 bp in length
* 46331 46430: gap of 100 bp
* 46431 47137: contig of 707 bp in length
* 47138 47237: gap of 100 bp
* 47238 47927: contig of 690 bp in length
* 47928 48027: gap of 100 bp
* 48028 48709: contig of 682 bp in length
* 48710 48809: gap of 100 bp
* 48810 49509: contig of 700 bp in length

| | | | |
|-------|--------|------------------|-----------|
| 33897 | 34683: | contlg of 787 bp | in length |
| 34684 | 34783: | gap of 100 bp | |
| 34784 | 35552: | contlg of 769 bp | in length |
| 35553 | 35622: | gap of 100 bp | |
| 35653 | 36406: | contlg of 754 bp | in length |
| 36407 | 36506: | gap of 100 bp | |
| 36507 | 37294: | contlg of 788 bp | in length |
| 37295 | 37394: | gap of 100 bp | |
| 37395 | 38159: | contlg of 765 bp | in length |
| 38160 | 38259: | gap of 100 bp | |
| 38260 | 39017: | contlg of 758 bp | in length |
| 39118 | 39117: | gap of 100 bp | |
| 39118 | 39890: | contlg of 773 bp | in length |
| 39891 | 39980: | gap of 100 bp | |
| 39991 | 40774: | contlg of 784 bp | in length |
| 40775 | 40874: | gap of 100 bp | |
| 40875 | 41643: | contlg of 769 bp | in length |
| 41744 | 41743: | gap of 100 bp | |
| 41744 | 42521: | contlg of 778 bp | in length |
| 42522 | 42621: | gap of 100 bp | |
| 42622 | 43393: | contlg of 772 bp | in length |
| 43394 | 43493: | gap of 100 bp | |
| 43494 | 44268: | contlg of 775 bp | in length |
| 44269 | 44368: | gap of 100 bp | |
| 44369 | 45149: | contlg of 781 bp | in length |
| 45150 | 45249: | gap of 100 bp | |
| 45250 | 46024: | contlg of 775 bp | in length |
| 46125 | 46124: | gap of 100 bp | |
| 46125 | 46688: | contlg of 744 bp | in length |
| 46689 | 46968: | gap of 100 bp | |
| 46969 | 47720: | contlg of 752 bp | in length |
| 47721 | 47820: | gap of 100 bp | |
| 47821 | 48506: | contlg of 776 bp | in length |
| 48597 | 48686: | gap of 100 bp | |
| 48697 | 49462: | contlg of 766 bp | in length |
| 49463 | 49562: | gap of 100 bp | |
| 49563 | 50335: | contlg of 773 bp | in length |
| 50336 | 50435: | gap of 100 bp | |
| 50436 | 51207: | contlg of 772 bp | in length |
| 51208 | 51397: | gap of 100 bp | |
| 51308 | 52078: | contlg of 771 bp | in length |
| 52079 | 52178: | gap of 100 bp | |
| 52179 | 52954: | contlg of 776 bp | in length |
| 52955 | 53054: | gap of 100 bp | |
| 53055 | 53813: | contlg of 759 bp | in length |
| 53814 | 53913: | gap of 100 bp | |
| 53914 | 54675: | contlg of 762 bp | in length |
| 54676 | 54775: | gap of 100 bp | |
| 54776 | 55537: | contlg of 762 bp | in length |
| 55538 | 55637: | gap of 100 bp | |
| 55638 | 56413: | contlg of 776 bp | in length |
| 56414 | 56513: | gap of 100 bp | |
| 56514 | 57278: | contlg of 765 bp | in length |
| 57279 | 57378: | gap of 100 bp | |
| 57379 | 58149: | contlg of 771 bp | in length |
| 58150 | 58249: | gap of 100 bp | |
| 58250 | 59023: | contlg of 774 bp | in length |
| 59024 | 59123: | gap of 100 bp | |
| 59124 | 59906: | contlg of 783 bp | in length |
| 59907 | 60006: | gap of 100 bp | |
| 60007 | 60792: | contlg of 786 bp | in length |
| 60793 | 60892: | gap of 100 bp | |
| 60893 | 61663: | contlg of 771 bp | in length |
| 61664 | 61763: | gap of 100 bp | |
| 61764 | 62539: | contlg of 776 bp | in length |
| 62540 | 62639: | gap of 100 bp | |
| 62640 | 63422: | contlg of 783 bp | in length |

| | | | | |
|-----------------------|--------------|--------------------|---------------|---------------|
| Query Match | 10.6%; | Score 443; | DB 64; | Length 72157; |
| Best Local Similarity | 98.9%; | Pred. No. 5.7e-81; | | |
| Matches 446; | Conservative | 0; | Mismatches 5; | Indels 0; |

QY 3624 gtctgaattggaagcctagagcgcaagaaaaaacgccgttcacagaacagggagaa 3683

| | | | |
|----|-------|--|-------|
| Db | 39686 | CTCTGAATTGGAGAAAGCCTAAGAGCGAGAAAAAAGCGCCGCTACAGACAGAGAGGAA | 39627 |
| Oy | 3684 | atlaagtaatggaatgacttgacttaagtttgtaacgaacagaacactlaagtcagtcag | 3743 |
| Db | 39626 | ATTAGTGATGATGATGACTTGACTTAAGTTGGTACAGAAACAGAAACCTTAAGCACTCAGCG | 39567 |
| Oy | 3744 | aagtcggaagaagaagggccatacgcgtctcagaatcttgatgaacagcagttgcttgagaaa | 3803 |
| Db | 39566 | AAGTCGGAAGAAAGAGCGCCCTACGCGCTTCAGAACTTGATGTAAACAGCACTGGCGCTTGAGGAAAA | 39507 |
| Oy | 3804 | gaggtcgaagaagaatatataatagaagaatgaaatgaaacagatatgttcgccaataaagg | 3863 |
| Db | 39506 | GAGGCTCAACAAAGATTAATTATTAAGAAATGAAGATGAACAGATATGTCGCCCAAAAAAGGG | 39447 |
| Oy | 3864 | taaaagagcgccagacacacaaacctcttgtytggaagtaacaaagaagaagccacaat | 3923 |
| Db | 39446 | TAAAAAGAGCGCCGACCACCAAAACCTCTTGTTGGTGAGGATACACCAAAAGAGGCCAACAT | 39387 |
| Oy | 3924 | gaaacctctlaaaaaaggaagcaaaaaaaatctcgtgacctcagcacagagagggaga | 3983 |
| Db | 39386 | GAAACCTTCTAAAAAAGAGAACCAAAAAAATAATCTGGAACCTCCACACCAAGAGGAGGA | 39327 |
| Oy | 3984 | agaagaagaagaacaaagtggaaatacoggaacagaagtccaagaacacagacagagt | 4043 |
| Db | 39326 | AGAAAGAAAGAAAGACAAAGTGGAAATTACGGAACAGAAAGTCCAAAAGCAACAGCACGAGT | 39267 |
| Oy | 4044 | gtcaagagagcacagcacaggaagcagaatct | 4074 |
| Db | 39266 | GTCAAGAGAGACACAGCGAGAGGTAAAGCTGT | 39236 |

| | | | | | |
|------------|----|--|--|--|---------|
| AC068224/c | 10 | | | | RESULTS |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| TITLE | | | | | |
| JOURNAL | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |

| | | | | |
|---|------------|-----|-----|-------------|
| AC068224 | 54398 bp | DNA | HMG | 30-APR-2000 |
| homo sapiens chromosome 3 clone RP11-660H19 map 3, | | | | |
| SEQUENCE SAMPLING. | | | | |
| AC068224 | | | | |
| AC068224.1 | GI:7671284 | | | |
| HGT: HTGS_PHASED. | | | | |
| human. | | | | |
| Homo sapiens | | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| 1 (bases 1 to 54398) | | | | |
| Birren,B., Linton,L., Nusbaum,C. and Lander,E. | | | | |
| Homo sapiens chromosome 3, clone RP11-660H19 | | | | |
| Unpublished | | | | |
| 2 (bases 1 to 54398) | | | | |
| Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barra,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Illey,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Larocque,K., Lamazares,R., Landers,T., Lebecky,J., Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Melidim,J., Meunier,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Murphy,T., Nayrou,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M., Oliver.J., Peterson,K., Pierre,N., Pisanal,C., Pollara,V., Raymond,C., Riley.R., Rogov,P., Rotlman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M. | | | | |
| Direct Submission | | | | |
| Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA | | | | |


```

Db      21073  TATATATTAATCATTAATTTGATCTATTCACATTTGGTCTCCCAACAGACAGACA 21132
Qy      4071  attcccggaatcagtcgaatgataccacagctccacaccagaagaagagag 4130
Db      21133  ATCCCGAATCTGATGATTAATTCACACAGCTCCACACAGAGAAGAGAGAG 21192
Qy      4131  accatcaaaaacgcacatcacacaacaaaataatgtg 4173
Db      21193  ACCATCAAAAAGCGCATTCACATCAACAAAAAATGTG 21235

RESULT  9
AC016449/c  AC016449  72157 bp  DNA  HTG  13-JUL-2000
LOCUS      Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC016449
ACCESSION  AC016449
VERSION    AC016449.2 GI:9104517
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 72157)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens, clone RP11-45L14
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 72157)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Batra,N., Beckerly,R., Boguslavsky,L., Bouknight,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearlano,R., Dewar,K., Domino,R., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Liu,C., Locke,K., MacDonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tsafte,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:6479175.
            All repeats were identified using RepeatMasker:
            smtc, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WtBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L1189
            Center clone name: 45_L_14

```

```

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

1 758 857: gap of 100 bp in length
* 858 1635: contig of 778 bp in length
* 1636 1735: gap of 100 bp
* 1736 2490: contig of 755 bp in length

```

```

* 2491 2590: gap of 100 bp
* 2591 3343: contig of 753 bp in length
* 3344 3443: gap of 100 bp
* 3444 4224: contig of 781 bp in length
* 4225 4324: gap of 100 bp
* 4325 5088: contig of 764 bp in length
* 5089 5188: gap of 100 bp
* 5189 5968: contig of 780 bp in length
* 5969 6068: gap of 100 bp
* 6069 6844: contig of 776 bp in length
* 6845 6944: gap of 100 bp
* 6945 7701: contig of 757 bp in length
* 7702 7801: gap of 100 bp
* 7802 8573: contig of 772 bp in length
* 8574 8673: gap of 100 bp
* 8674 9430: contig of 757 bp in length
* 9431 9530: gap of 100 bp
* 9531 10295: contig of 765 bp in length
* 10296 10395: gap of 100 bp
* 10396 11139: contig of 744 bp in length
* 11140 11239: gap of 100 bp
* 11240 11988: contig of 749 bp in length
* 11989 12088: gap of 100 bp
* 12089 12874: contig of 786 bp in length
* 12875 12974: gap of 100 bp
* 12975 13750: contig of 776 bp in length
* 13751 13850: gap of 100 bp
* 13851 14638: contig of 788 bp in length
* 14639 14738: gap of 100 bp
* 14739 15520: contig of 782 bp in length
* 15521 15620: gap of 100 bp
* 15621 16381: contig of 761 bp in length
* 16382 16481: gap of 100 bp
* 16482 17248: contig of 767 bp in length
* 17249 17346: gap of 100 bp
* 17349 18123: contig of 775 bp in length
* 18124 18223: gap of 100 bp
* 18224 18977: contig of 754 bp in length
* 18978 19077: gap of 100 bp
* 19078 19860: contig of 783 bp in length
* 19861 19960: gap of 100 bp
* 19961 20732: contig of 772 bp in length
* 20733 20833: gap of 100 bp
* 20833 21590: contig of 758 bp in length
* 21591 21690: gap of 100 bp
* 21691 22446: contig of 756 bp in length
* 22447 22546: gap of 100 bp
* 22547 23333: contig of 787 bp in length
* 23334 23433: gap of 100 bp
* 23434 24212: contig of 779 bp in length
* 24213 24312: gap of 100 bp
* 24313 25093: contig of 781 bp in length
* 25094 25193: gap of 100 bp
* 25194 25958: contig of 765 bp in length
* 25959 26058: gap of 100 bp
* 26059 26839: contig of 781 bp in length
* 26840 26939: gap of 100 bp
* 26940 27705: contig of 766 bp in length
* 27706 27805: gap of 100 bp
* 27806 28568: contig of 763 bp in length
* 28569 28668: gap of 100 bp
* 28669 29419: contig of 751 bp in length
* 29420 29519: gap of 100 bp
* 29520 30311: contig of 792 bp in length
* 30312 30411: gap of 100 bp
* 30412 31202: contig of 791 bp in length
* 31203 31302: gap of 100 bp
* 31303 32082: contig of 780 bp in length
* 32083 32182: gap of 100 bp
* 32183 32934: contig of 752 bp in length
* 32935 33034: gap of 100 bp
* 33035 33796: contig of 762 bp in length
* 33797 33896: gap of 100 bp

```


repeat_region
5620..5810
/note="MIR repeat: matches 58..262 of consensus"
7829..8133
/note="AluX repeat: matches 1..301 of consensus"
repeat_region
10957..11242
/note="AluY repeat: matches 291..1 of consensus"
repeat_region
12344..12405
/note="L1PA16 repeat: matches 903..842 of consensus"
repeat_region
12381..13119
/note="L1PA16 repeat: matches 833..68 of consensus"
repeat_region
14428..14540
/note="HY1 repeat: matches 110..1 of consensus"
repeat_region
14951..15071
/note="MIR2 repeat: matches 126..1 of consensus"
repeat_region
16589..16885
/note="AluX repeat: matches 302..4 of consensus"
repeat_region
18067..18397
/note="L1 repeat: matches 3807..4127 of consensus"
repeat_region
19076..19481
/note="MLT1A2 repeat: matches 374..1 of consensus"
repeat_region
19516..19799
/note="AluX repeat: matches 19..302 of consensus;
incomplete repeat"
repeat_region
19808..20127
/note="AluY repeat: matches 2..299 of consensus"
prim_transcript
21359..26899
/note="match: multiple ESTs; match: H44781 D20082 W20070
T64867 H45693; match: F12354 N91467 N83765 H45741 H93424"
repeat_region
24795..25089
/note="AluX repeat: matches 294..1 of consensus"
repeat_region
26083..26106
/note="L12 copies of 2 mer 96 & conserved"
prim_transcript
complement(27120..27476)
/note="match: multiple ESTs; match F09972 T72607"
prim_transcript
27633..28101
/note="match: multiple ESTs; match: 5' EST T51695 clone
72405; Paired with EST T51543 matching this clone; match
H91918"
repeat_region
28327..28360
/note="17 copies of 2 mer 88 & conserved"
prim_transcript
28864..29013
/note="match: 3' EST T51543 clone 72405; Paired with EST
T51695 matching this clone"
repeat_region
29300..29456
/note="MIR2 repeat: matches 146..1 of consensus"
repeat_region
30648..30934
/note="AluX repeat: matches 7..302 of consensus"
repeat_region
33156..33439
/note="AluX repeat: matches 301..2 of consensus"
repeat_region
34900..35062
/note="L1MD2 repeat: matches 425..263 of consensus"
repeat_region
34982..35061
/note="L1ME3 repeat: matches 343..264 of consensus"
repeat_region
37923..38245
/note="AluX repeat: matches 294..1 of consensus"
repeat_region
38457..38556
/note="MIR repeat: matches 61..161 of consensus"
repeat_region
39265..39568
/note="AluO repeat: matches 302..3 of consensus"
repeat_region
40952..41046
/note="MIR repeat: matches 29..125 of consensus"
repeat_region
43042..43331
/note="AluO repeat: matches 301..11 of consensus"
repeat_region
43583..43669
/note="MERSA repeat: matches 10..99 of consensus"
repeat_region
43881..44083
/note="MIR repeat: matches 248..28 of consensus"
repeat_region
44772..45170
/note="MIR repeat: matches 388..1 of consensus"
repeat_region
45848..45949
/note="MIR repeat: matches 144..41 of consensus"
repeat_region
46073..46373

repeat_region
49357..49436
/note="AluSp repeat: matches 298..1 of consensus"
repeat_region
50023..50324
/note="MIR repeat: matches 198..114 of consensus"
repeat_region
50826..50892
/note="AluX repeat: matches 1..302 of consensus"
repeat_region
51407..51534
/note="L1MB8 repeat: matches 853..921 of consensus"
repeat_region
51563..51863
/note="L1 repeat: matches 4869..4999 of consensus"
repeat_region
51869..52077
/note="AluY repeat: matches 1..301 of consensus"
repeat_region
52311..52712
/note="AluO repeat: matches 86..298 of consensus;
incomplete repeat"
repeat_region
52698..52781
/note="L1ME3A repeat: matches 44..475 of consensus"
repeat_region
52719..52833
/note="L1MD2 repeat: matches 569..653 of consensus"
repeat_region
53752..54490
/note="L1MB3 repeat: matches 591..707 of consensus"
repeat_region
54313..54870
/note="L1 repeat: matches 4668..5390 of consensus"
repeat_region
55008..55520
/note="L1MB4 repeat: matches 1..540 of consensus"
repeat_region
57267..57487
/note="L1MA9 repeat: matches 530..1055 of consensus"
repeat_region
57510..57576
/note="AluY repeat: matches 79..298 of consensus;
incomplete repeat"
repeat_region
58670..58699
/note="MIR repeat: matches 76..143 of consensus"
repeat_region
58676..58699
/note="L15 copies of 2 mer 87 & conserved"
repeat_region
59808..60115
/note="6 copies of 4 mer 100 & conserved"
repeat_region
60135..60526
/note="L1MB7 repeat: matches 895..571 of consensus"
repeat_region
60531..60815
/note="MIR1D repeat: matches 505..108 of consensus"
repeat_region
60816..60915
/note="AluO repeat: matches 2..290 of consensus"
repeat_region
60918..61113
/note="MIR1D repeat: matches 114..3 of consensus"
repeat_region
61115..61379
/note="L1MB6 repeat: matches 529..320 of consensus"
repeat_region
61380..61675
/note="AluY repeat: matches 299..31 of consensus;
incomplete repeat"
repeat_region
61540..61805
/note="L1MB5 repeat: matches 309..1 of consensus"
repeat_region
61916..62212
/note="L1 repeat: matches 5389..5107 of consensus"
repeat_region
64361..64420
/note="AluSc repeat: matches 1..299 of consensus"
repeat_region
64816..65101
/note="U6 repeat: matches 63..1 of consensus"
repeat_region
65202..65856
/note="AluO repeat: matches 2..302 of consensus"
repeat_region
66393..66470
/note="L1PA16 repeat: matches 902..217 of consensus"
repeat_region
66530..66868
/note="L1 repeat: matches 2101..2178 of consensus"
repeat_region
66881..67104
/note="L1 repeat: matches 2265..2610 of consensus"
repeat_region
67104..67855
/note="MERS30 repeat: matches 2..230 of consensus"
repeat_region
67999..68677
/note="L1 repeat: matches 2613..3397 of consensus"
repeat_region
68693..68972
/note="L1 repeat: matches 4255..4942 of consensus"
repeat_region
68983..69436
/note="AluSc repeat: matches 299..1 of consensus"


```

REFERENCE
AUTHORS      1 (sites)
              Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
              Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
              Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE        NEO human cDNA sequencing project
JOURNAL      unpublished (2000)
REFERENCE    2 (bases 1 to 2079)
AUTHORS      Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
              Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE        Direct Submission
JOURNAL      Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
              Sugano, Institute of Medical Science, University of Tokyo,
              Laboratory of Genome Structure Analysis, Human Genome Center;
              Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
              (E-mail:cdna@ems.u-tokyo.ac.jp, Tel:81-3-5449-5286,
              Fax:81-3-5449-5416)
COMMENT      NEO human cDNA sequencing project supported by Ministry of
              International Trade and Industry of Japan; cDNA full insert
              sequencing: Research Association for Biotechnology; cDNA library
              construction, 5'- & 3'-end one pass sequencing: Department of
              Virology and Human Genome Center, Institute of Medical Science,
              University of Tokyo (partly supported by science and Technology
              Agency).
FEATURES
Source       Location/Qualifiers
             1..2079
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="COL00725"
             /clone_1fb="COL"
             /issue_type="colon"
             /note="Cloning vector pME18SFL3"
             304..681
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="BAB15584.1"
             /db_xref="GI:10439855"
             /translation="MDLJTLVQEQPKGSRKRRKHTASEDSQVPEEKRLKEDI
             LENEDEONSPKGRGRRPKPLGCGTKEPTMTSKSKSKSGPPAPEEEDEER
             OSGNTEQSKSKKHVSRRQOR"
CDS
             759 a      354 c      339 g      627 t
BASE COUNT
ORIGIN
Query Match      10.7%; Score 447; DB 89; Length 2079;
Best Local Similarity 85.5%; Pred. NO. 8.1e-82;
Matches 550; Conservative 0; Mismatches 0; Indels 93; Gaps 1:
OY 3624 gctgaattggagaagccttagagcaggaagaaacgcgcgtcacagaacagagagaga 3683
    |||||||
DB 237 gctgaattggagaagccttagagcaggaagaaacgcgcgtcacagaacagagagaga 296
OY 3684 attagatgagtagcttactcaagttgttacagagacagaacctaagaagcagtcagcg 3743
    |||||||
DB 297 attagatgagtagcttactcaagttgttacagagacagaacctaagaagcagtcagcg 356
OY 3744 aagtcggaagagagccatagcgcctcagaatctgtagaacagcagtcgagtcgaagaaa 3803
    |||||||
DB 357 aagtcggaagagagccatagcgcctcagaatctgtagaacagcagtcgagtcgaagaaa 416
OY 3804 gaggtctcaagaagatattatagaataatgaatgaacagaatagtcgcgcaaaaaaagg 3863
    |||||||
DB 417 gaggtctcaagaagatattatagaataatgaatgaacagaatagtcgcgcaaaaaaagg 476
OY 3864 taaagaagcgcagcaccacaaacctctgtgtgaggtacacacaaagaagagcacaat 3923
    |||||||
DB 477 taaagaagcgcagcaccacaaacctctgtgtgaggtacacacaaagaagagcacaat 536
OY 3924 gaaaccttcaaaaaagaagcaaaaaaatctgtgacctcagcaccagagagagaga 3983
    |||||||
DB 537 gaaaccttcaaaaaagaagcaaaaaaatctgtgacctcagcaccagagagagaga 596
OY 3984 agaaagaagaagaagaagagagagagagagagagagagagagagagagagagagagag 4043
    |||||||

```

```

DB 597 AGAAGAAGAAAGCAAAAGTGGAAATAGCGAAGACGAAGTCCAAAAGCAACAGCAGCAGT 656
OY 4044 gtcaagaagagcagcagcaag-----
    |||||||
DB 657 GTCAGAGAGAGCAGCAGCAGAGGTAAGCATGTGTACTTAACCTGCATCTGTTCTTAC 716
OY 4063 -----
    |||||||
DB 717 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 776
OY 4071 attctctgaattagtgcaatgtgaatcacaacagcagtcacacacagaagaagagagag 4130
    |||||||
DB 777 ATCTCTGTAATCTATAGCAATTTGAATTCACACAGTCCACACACAAAGAGAGAGAG 836
OY 4131 accatcaaaaacgcacatcacatcacacacacacacacacacacacacacacacacacac 4173
    |||||||
DB 837 ACCATCAAAAAGCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 879
RESULT 7
HS49J10 137246 bp DNA PRI 22-NOV-1999
LOCUS
DEFINITION Human DNA sequence from PAC 49J10, BRCA2 gene region chromosome
13q12-13 contains ESTs.
ACCESSION 284572
VERSION 284572.1 GI:1813974
KEYWORDS 13q12-13.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 137246)
REFERENCE
AUTHORS Whiteley,M.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1HQ, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
IMPORTANT: This sequence is not the entire insert of clone . It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 49J10 is at 1 in this sequence. The true
right end of clone 267P19 is at 43908.
The true left end of clone 179J15 is at 137143.
49J10 is from the human PAC library described in Ioannou A.P. et al
Nature Genet 6, 84-89.
FEATURES
Source       Location/Qualifiers
             1..137246
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="13"
             /map="13q12-13"
             /clone="XX-49J10"
             1531..1831
             /note="AluX repeat: matches 281. .1 of consensus;
             incomplete repeat"
             repeat_region
             1838..1871
             /note="17 copies of 2 mer 100 % conserved"
             repeat_region
             1882..2741
             /note="L1P1.5 repeat: matches 893. .1 of consensus"
             repeat_region
             2597..2848
             /note="L1 repeat: matches 5389. .5139 of consensus"
             repeat_region
             3149..3405
             /note="MER44A repeat: matches 247. .1 of consensus"
             repeat_region
             3853..3922
             /note="AluSc repeat: matches 188. .120 of consensus"

```


Db 121 CTACATCAGAGGCTTACTGTTCTGCCATGTTTGAAAACTGATGACCATACAGAAAGATT 180
 QY 1607 taccgtgacctgtaagagctcaagattctcgaagaatcacacagtgctgaagaag 1666
 Db 181 TGCGTGACCCCGGGAAGSCACAGATTGTTGGAAGAAATTTAACACAGGTTCTCGCGCATG 240
 QY 1667 atgagaataaagaagcagtlagaagtlactgtltagtccacatgctcctcagcag 1726
 Db 241 ATGAGAAATCGGCTCAGTTGAGTTGATTAATTAGCCCAACCTGTTCTTGCAACAG 300
 QY 1727 ctgaagtgctgctggaataactaagaagtggtgacaccccaacagcctacaac 1786
 Db 301 CAGATATTTGTTGAGAGAAATACCCGGAACCTTGCAAAATCTTGAACCAACCAATTC 360
 QY 1787 ctcccggaatgatcaagttctctctggaagatagacacctggtacatagatagc 1846
 Db 361 CTGTTCTAGAGATGCTCAAAATTTCTGTTGGAAGAAATCGACCTGTGACACTTGTATG 420
 QY 1847 aatcatalcagtgctctattataaacaagtgaaacaatatagatgaaacagcagatg 1906
 Db 421 AAGCATTAAGTGCACCTAGTGAATTTGATGAATGATAGTCAATAGAGGGGACAGCATG 480
 QY 1907 aagtgagtggtgtcccaatgataagcaccagagctcctggaactgttaagtaac 1966
 Db 481 AAGGAGGAGGTGATGTCACATACAGCTATCCGTTCAAGACTTGAACCTTCTTAAGTTTC 540
 QY 1967 tctcattacacatcccatccatccattcctgctgtaagaactggaatccatctagctt 2026
 Db 541 TGCTTTTACACATCTTACCTCGTTCCACTGTGACAGACATATGAGTCTTGTATGACT 600
 QY 2027 gctcgaataatgaaatgaaagtagaagaagctgacatacaatctcacaacacag 2086
 Db 601 GCCTAAATAATGAGATGACAGAGTAGAGAGAGCTGATTAATTTTGAATAATACAG 660
 QY 2087 gaagcaaatatgaagaagatttccacacatcagatcagcctgctcctccttcaatc 2146
 Db 661 GTCCAAATAATGAAACAGACCTTCCCAAGATGATGATGATGATGATGATGATGATG 720
 QY 2147 acaatctaaagaag 2206
 Db 721 AAAAAGCAAGAGAGGATGATCTCCACACAGCAAGCAAGAGAGAGAGAGAGAGAGAG 780
 QY 2207 taattctagaagaag 2266
 Db 781 TATTCACAAATAATGAAAGATGACAGCTGATGATGATGATGATGATGATGATGATG 840
 QY 2267 atccaagaacactggaacatctcaacacatcagatggttactatggtatctgctccc 2326
 Db 841 ATGCTGATGTCAGAAACAACTTATTAATCTCATTAGTTTCAATGGGCCCAATTTCTATGT 900
 QY 2327 ttgcaactgatacaattgctgctccttgaagaatcttggtagtacttacttctgaaga 2386
 Db 901 TAGACACAGATCACTTTGCTTCCCAATGAAATCTGTATGCAAAATTTTATGTGAAG 960
 QY 2387 atctctcatgaatgatacgtgctcctcaggaagaagaacaactaaacttgggtlccaga 2446
 Db 961 ATCTGTAATGATGAACAGCTCAACAGGTGAAAGAAATGAAACTGTGCTCCAGATG 1020
 QY 2447 aagaagatctcctcaggaagaatgltcaaaatcagcgtcattataaagtagtctgagc 2506
 Db 1021 AAGAGGTTTCCCTGAACTACTACAAAGTACAGGCAATTAATCTGTAAGAGTGCC 1080
 QY 2507 tactggaatgaaataatcacacagtaaatcaggaactctacacttaagatgctgaaca 2566
 Db 1081 TGTTGGGTATCAAAAACACAGCTTAATCTGCAATTTCAACCTTGGTTATATATAG 1140
 QY 2567 caatattgcatgtagtgaagactgacagaacaggggaaataatgataaacagatatgt 2626
 Db 1141 CGAGTGTGTTAGTAGAGGTGACCTGACAGCAAAAGAGATGATTAATCTGATATGT 1200
 QY 2627 caagtcgagagactgctgctggtgaggtctatctggaagcctgacagaacacccgttacc 2686
 Db 1201 CTCGCTTGCGATTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260

QY 2687 atgaatcatcacatlaagaacatalcagctatgcatlaagatatacaacgatgaatgt 2746
 Db 1261 ATGAAATATTATACCCAGACAAACAGTTTACGCTGTGCGACTGTATTAATATGATGATGT 1320
 QY 2747 atcaagtaagaagaagtggtcccaagaactcacaagaagccttccgtttagcgttc 2806
 Db 1321 ACCAAGTAAAGCAGATATTCTGCAAGAGCTGCAATTAAGCACTGTGAAAGTACGCTCC 1380
 QY 2807 cacttgagtatatgcaatctgctgctccttctgcaaaaagatccctglaaagaagaag 2866
 Db 1381 CATTGAGTATATGCGCGATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
 QY 2867 ctcatcgaagcaatglttggtagaanaataatgtaagcgggaagatctggaagcagc 2926
 Db 1441 CACACGACGACATGTTTACTGAAATATACAGATATGCGAGGAAATACATTAAGCAGA 1500
 QY 2927 atgaagctgttagtgaanaaatatctctcctcaccagatgltgttccatatcaa 2986
 Db 1501 ATCCATGCTACTGAGAAATTAATATATATATATATATATATATATATATATATATAT 1560
 QY 2987 ttcacacttggcaatgacccagatattgccaagatcagatattgaacaacttaag 3046
 Db 1561 TTCACCTGCTAGCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
 QY 3047 atgttaagaatgltctgt 3106
 Db 1621 ATATCAAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
 QY 3107 gtcacgctttatcagaagaatgtagaanaatatataaacaacaagaatgcccagagac 3166
 Db 1681 GCCATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
 QY 3167 cagatgtagcaaaaatgatatataaactgtacactgtgtgtgtgtgtgtgtgtgtgtgt 3226
 Db 1741 CAGATGAATCCAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800
 QY 3227 tcatgtaagaagatactacatacaagtlggaatctcctcaagaagcctgactaccagctc 3286
 Db 1801 TAAATATGTAAGATGCTTGTGCAATGCAAGATGCAACCAAGAGACCATCTCCCAATG 1860
 QY 3287 gttccttcaactcaactgcaagaatctcagtaacaacacaataatctgcctcctgaa 3346
 Db 1861 AATTTTATACCAACCTGTAAGAGACTTCTGTATGATGATGATGATGATGATGATGATG 1920
 QY 3347 tgaatcatcttcaactcctggaagaacctaagaacaaagctgtctgagagctgttaca 3406
 Db 1921 CAAAGTACTTCTGTTAAGAGAAAGCCAAAGCCTGCTGAGTACTAGTCCAGTAATTA 1980
 QY 3407 agccaacttcatcagcagcagcaatctcagacccaatctcaacgaatggaatgaaactgaa 3466
 Db 1981 AGCCTTATCAGCAAGGGAAGAAACCTTATGTAAGAGCACTGCACTGCACTGCA 2040
 QY 3467 gcaatgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3526
 Db 2041 GCAATATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100
 QY 3527 gttcgaatgtagatcacagtgtaaaatgaga 3587
 Db 2101 GTTCAGAGCAGCAAGAACTGAGTTACTGA 2131

RESULT 6
 LOCUS AK026889 2079 bp mRNA PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLJ23236 f1s, clone COL00725.
 ACCESSION AK026889
 VERSION AK026889.1 GI:10439854
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens colon cDNA to mRNA, clone lib:COL clone:COL00725.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

| | | | | | | |
|------------|--------------|---|-----------------------------------|-----------------------|-------------|--|
| Db | 841 | ATGAAGATGAGGGGTTC | CAAAGTCAATCAACCATTCAGACAGGTCTTGAC | CTTAAG | 900 | |
| OY | 1964 | tacttccatttcaacatcccatctcatcttcaatcttgc | tgcgtaaacatttgaatcaattacttgc | 20232 | | |
| Db | 901 | TACTCATATTACACATCCCATCTCATTTTCATTTCT | CTGTAACAACTTTGATCATTTACTG | 960 | | |
| OY | 2024 | ctttcttgcgaanaatgagatgagaaagaagcagaagc | cttccctacaatcttcaaaaaca | 20893 | | |
| Db | 961 | CTTGCTGTGAANAATGATGATGAAAAGTACACAAAG | CTCCTACAAATTTTCAAAAACA | 10200 | | |
| OY | 2084 | caggaaagcaaaatltgaagagagatcttccacacatcc | agatccagccttgccttccgttttac | 21433 | | |
| Db | 1021 | CAGGAGCAAAATTTGAAGAGATTTTTCACACATCAG | ATCAGACCTTGCTTCCTGTTTAC | 10880 | | |
| OY | 2144 | atcacaaatcttaaaaaagagcccccccg | tcgaagccaataatgccaattcatgtatccat | 22022 | | |
| Db | 1081 | ATCCAAATCTAAAAAAGAGACCCCCCGCTCAAGCC | AAATATGCCAATTCATTTGATTCAT | 11400 | | |
| OY | 2203 | ggcgaattttcttagtaaaagaccacagtttgcacagat | atttgagccttgcataaagac | 22623 | | |
| Db | 1141 | GGGATATTTTCTAATTAAGAGACATCGATTGTGC | ACATATTTTGACCTCTCGATTAAGAC | 12000 | | |
| OY | 2263 | ctagatcccaagcaaccttgcacatctcataaaccaat | tggttactatitgltcatatgct | 23222 | | |
| Db | 1201 | CTAATCTCCAAAGCAACCTGGAAACATCTCATACACC | ATTGTTGTTACTAATTTGTCATTTGCT | 12600 | | |
| OY | 2323 | ctctcttgcacctgtatcaatttgccttgccttgcgaat | ctcttggtgtgccttacttcaatttgc | 23822 | | |
| Db | 1261 | CTCCTTGACCTGTACATTTTGCTGCTCTTGAAATCT | TTGTAAGCTTCACTTCAATTTG | 13200 | | |
| OY | 2383 | aaagaatcttctcatgaaatgtgcgcttccaggggaaa | agaacacaacttcaacttgggttctca | 24422 | | |
| Db | 1321 | AAAGATCTTTCATGATGATGATGCGCTTCCAGG | AAAAAGACAATCAATCAATTTGGTGTCA | 13800 | | |
| OY | 2443 | gataagaagaatcttccctgtgcgaacaaatggttcaaa | aatctcagcctcataaataatgtatgtctga | 25022 | | |
| Db | 1381 | GATTAAGAAGATCTCTCTGGAACAATGTGTAATAAT | TTCAGGCTATTAAATGTATGGTGTGA | 14400 | | |
| OY | 2503 | tggtctacttgcgaatgaaanaatactcaagtaaatcca | gagaaacttcaacttaag | 25612 | | |
| Db | 1441 | TGGCTACTTGGCATGTAAATTTAATTCACAGTATAT | TCAGGATCTTCTCTTAAGSATTTGCT | 15000 | | |
| OY | 2562 | aacaacataatgtcatagatgtatgagaaacttgcaga | acacaggggaaanaatagtaaacaga | 26212 | | |
| Db | 1501 | AACAACATATTCGATCGTATGATGGGA | -TTGMAAGAMCAGGGGAAA- | TTACTATACAGA | 15572 | |
| OY | 2622 | tatttcagcttgcgaacttgcctgtcgagagtgatatt | gtgaagcctgcgcacaagaacctgc | 26812 | | |
| Db | 1558 | TATGTCACGCTGTGAGCTTCTCTCTGTGGAGTGTAT | TTTGAAGCTGTGGCAAGAAGCCCTGC | 16172 | | |
| OY | 2682 | ttaccatgaaatactcatcatctagaacaatactcagc | catatgtcatatagcatatcaacagatga | 27412 | | |
| Db | 1618 | TTACCATGGAATCTCTCATMTAGAACATATCAGTAT | GCTTATGCTTTAGCTATCAACAGATGA | 16772 | | |
| OY | 2742 | atgtcatcaagtaagaacaagtgtttc | gccacagaacctcaacaaggccttccggtttac | 28000 | | |
| Db | 1678 | ATGCTATCAAGTAAAGACAAATGTTTGGCCAGAA | ACTTGTGGCAAAAGCCCTTCCCTTTGTC | 17372 | | |
| OY | 2801 | ggcttccaaacttgcgtatataatgcaactgttgcctt | gtgtcaaaaagatccgttgaagaagga | 28650 | | |
| Db | 1738 | GGCTTCACCTTGAAGAATATGTGCAATCTGTG | CCCTTTGTGTCAAAAAGATCTCTGTAAAGAGGA | 17972 | | |
| OY | 2861 | gaagagactcatgtctaggcaatgtttgtgtgataaaa | tataatgataatgtaagcggagagt | 2914 | | |
| Db | 1798 | GAAAGCTCATGCTAGGCAATGTTTGGTGA | AAAAATTTAATGTAAAGCGGGAGCT | 1851 | | |
| RESULT | 5 | | | | | |
| AB014548 | | | | | | |
| LOCUS | AB014548 | 5177 | bp | mrna | PRI | |
| DEFINITION | Homo sapiens | mrna | for KIA00648 | protein, partial cds. | 06-FEB-1999 | |
| ACCESSION | AB014548 | | | | | |

| VERSION | AB014548.1 | GI:3327109 |
|----------------------------|---|------------|
| KEYWORDS | Homo sapiens adult male brain cDNA to mRNA, clone_lib:pb1uescript11 | |
| SOURCE | SK plus clone:Hu03994. | |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | |
| AUTHORS | Obara,O., Suyama,M., Nagase,T. and Ishikawa,K. | |
| TITLE | 1 (bases 1 to 5177) | |
| JOURNAL | Direct Submission | |
| | Submitted (25-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu Obara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yama 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@fokkazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914) | |
| REFERENCE | 2 (sties) | |
| AUTHORS | Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Obara,O. | |
| TITLE | Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro | |
| JOURNAL | DNA Res. 5 (3), 169-176 (1998) | |
| MEDLINE | 98403880 | |
| FEATURES | Location/Qualifiers | |
| source | 1..5177 | |
| | /organism="Homo sapiens" | |
| | /db_xref="taxon:9606" | |
| | /clone="Hu03994" | |
| | /clone_lib="pb1uescript11 SK plus" | |
| | /dev_stage="adult" | |
| | /sex="male" | |
| | /tissue_type="brain" | |
| gene | 1..2558 | |
| | /gene="K1AA0648" | |
| | <1..2558 | |
| | /gene="K1AA0648" | |
| | /codon_start=3 | |
| | /product="K1AA0648 protein" | |
| | /protein_id="BAA31623.1" | |
| | /db_xref="GI:3327110" | |
| CDS | /translation="IYYIVASLDIPNAVKALNEMMKCOMLRSVHREILDIHKOPSE
NCSAMFKLMTAKNLDPGKAODEPVKFNVLVLDDEKLRSKLELLISPTSCQAD
CVRIRAKRLANPKQPTNPFLELVKFLERIAVHIDSAISALVKLNMKSTEGRADDE
BEQVSPTDAIRSGLELTVLSTHTPTSHSAETYESLQCRIMEDKVAEAA10IFR
TGHLETDLPQIRSTLLPIHOKAKRGPHOKAVH1A1FTNKEVOLAIEPLS
RSLNADVEOELITPVLVSGHISMLADPOFASPMKSVANFTVKDLMNDSTGEKNG
LMSDEEVSPEVLAKVQA1KILVRLMGLMKRKNOSKSNASTRLISAMVSGDLTLEOK
RISKSDMSRLAAGSAIMKLAQEPCHYEITTPPOCALVINDECVQVQITRAKRI
HSAVLKLLPLELMAIFALCAKDPDKERRAAROCCLANISIREYIKONPWATEKL
SLBEVYVPMVHILHLADPDRSQDVALDIECLMEFLVLEMTKKNNSHAFMK
MAENIKLITRDAOSDESKTNEKLYTVDVALCVINSKSLNADSPKDPVLPKPFQ
PEKPCNDKSYISEETRYLLTGKPKPAGVGLAVNPLSATGRKPYVSTGETGNSN
NVNSELNSTGNSREOSESAAETGVSENEENPVRIISVTPVKNIDPVKNKINSDOA
TQGISDSDRGKRTVTAAGAENIQOKDEKVEDSGPAPSRBRGRPKSRSGSNATW
NDNLNKPINKRRKRAAYQGESPGCLEAGNAKAPRLQDLAKAAAEAOIDQR" | |
| BASE COUNT | 1690 a 931 c 1037 g 1519 t | |
| ORIGIN | | |
| Query Match | 25.0%; Score 1041.4; DB 85; Length 5177; | |
| Best Local Similarity | 68.0%; Pred. No. 4.9e-204; | |
| Matches 1450; Conservative | 0; Mismatches 681; Indels 0; Gaps 0; | |
| OY 1427 | gctatatactgtatgcacacactgattaaatgctgtgtgaagaacatgaatgaatgt 1486 | |
| | | |
| Db 1 | GGTTATTATTACTTAATGCTAGTTGGATCCAAATGCTGTAAACCTCTCAACGAATGT 60 | |
| | | |
| OY 1487 | ggaaatgtcaaatgctgcgcgaacatcaagtaagaagattgcttgaactgattgaagcaac 1546 | |
| | | |
| Db 61 | GGAAGTGCAGAACATGCTTCGAGCCATGTACGGGAACATTGGATTTCGACAAAGCAC 120 | |
| | | |
| OY 1547 | ccaaaacgatgtccagtgtaacagcatatttccaagaatgtgttalttaacgaanaatt 1606 | |
| | | |

1412 cagaacgagatgaatgactatatactctglatgacacactgtaattaaatgctgtaag 1471
1321 CAGAAGGATGAAATGCTATATTAATCTGTATGCCACATGATTTAAATGCTGAAG 1380
1472 ccttgaatgaatgtaagaaatgtaaaatctgctccgacataaagtaagatctgtg 1531
1381 CATTGAATGAATGTAAGTAATGCAAAATCTCGACATCAAGTAAGATTTGCTTG 1440
1532 acttgaatgaacacccaacagatgctcagctgtaagccatatttcaaaatgtaag 1591
1441 ACTTGATTAAGCAACCCAAACAGATGCCAGTCTCAAGCCATATTTTAAAAAGTATG 1500
1592 ttattacaagaattaccctgaacctgtaaggtcgaaggtcgaaggttcaatgaatcaac 1651
1501 TTATTACAAGAATTATTAATCTGATCTGTAAGGCTGAGATTTTCAATGAATTCACAC 1560
1652 aggtgttgaagaatgtaagaaatgaagaagcagttgaagatctgttgaatcaacat 1711
1561 AGGTGTGAAGAATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1620
1712 gctcctgcaagcagctgaaggtgtgtgctgtaaatgaatgaaggttggcaaccca 1771
1621 GCTCCTGCAAGCAGGCTGAAGTGTGTGCTGTAAGTAAGTAAGTAAGTAAGTAAG 1680
1772 aacagcctacaaatcccttctcgtgaatgaatgaatgttcttggagagatagaacctg 1831
1681 AACACCTCAAAATCTTCTCTGTAAGTATGATCAAGTTTCTTGGAGATAGCACCTG 1740
1832 tgcacatagatccgaatcactacagctcttattaaagaatgaagaatgaatgaatg 1891
1741 TGCACATGATCCCAATCTATCAGTCTTATTAAGTAAGTAAGTAAGTAAGTAAG 1800
1892 gaacagcagatgaatgaatgaatgaggtgtgtcacaatgaatgaatgaatgaatg 1951
1801 GAACAGCAGATGATGAAGTGAAGGTGTTCACAGTATCAACCAACAGCAGCTTG 1860
1952 aactgcttaagtaactccttcaatcaatcaatcaatcaatcaatcaatcaatcaat 2011
1861 AACTGCTTAAGTACTCTCATTTACATCCATCTCATTTCTGCTGAACATTTG 1920
2012 aacttactgctgtctgtaaatgtaagtaagtaagtaagtaagtaagtaagtaagta 2071
1921 AATCAATTAATGCTGTGCTGAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1980
2072 ttcttcaaaacacgaagaacaaatgtaagagatcttccacacatcaatcaatcaat 2131
1981 TTTTCAAAAACACGAAGCAAAATGTAAGAGATTTTCCACACATCAATCAGCTTGC 2040
2132 ttcctgtttacatcaaaatctaaanaagaccccccgtaagccaataatgcatc 2191
2041 TTCCGTGTTTACATCAAAATCTAAAAAAGGACCCCCCGTCAAGCAAAATATGCCATTC 2100
2192 atgtatcaatgagatatttctagtaagagacccaatgtaagatattgagccctc 2251
2101 ATTGATTCATGCGATATTTCTAGTAAGAGACCCAGTTTCCACAGATTTTGAACCTTC 2160
2252 tgcataaagccttagatccaaacacactggaacatctcaataacacatgtaactatg 2311
2161 TGCATAAGAGCCTTAGATCCAAACCACTGGAACATCTCTAACAACCATTTGTTACTATTG 2220
2312 gtcataatgctcctcctgcaactgtaacaaatgtgtgctccttggaaatcttggtagcta 2371
2221 GTCATAATGCTCCTCTGCACTGATCAATTTGCTGCTCTTTGAAGATTTTGGTAGTA 2280
2372 ccttcaatgtaagaatgcttccatgaatgatacgtcttcgaaggaagaagaacacaaac 2431
2281 CTTTATTTGTGAAGAATCTTCTCATGATGATGCGCTTCCAGGAAAGAAACAACTAAAC 2340
2432 ttctgggttccagaatgaagaatgatacctcctggaacaaatggttcaaaatctcaggtlat 2491
2341 TTTGGGTTCCAGATGAAGAATGATCTCTGAGACAAATGTTCAAAATTCAGGCTATTTAAA 2400
2492 tgatggttcgatgctacttggaaatgaaaaaataatcaatcaatcaatcaatcaatc 2551

2401 TCATGTTTCGATGCGCTACTTGAATGAATAAATACAGTAAATCAGGAACCTTACT 2460
2552 taagattgctaaacaacaaatattgcatagatgagagacttgcacaaacggggaataa 2611
2461 TAAGATTGCTAAACAACAAATTTGATGATGAGACTTGCACAAACGGGGAATAA 2520
2612 gtaaaccaagatattgcaactcgaagacttgcctcgggagatgtaattgtaagctgac 2671
2521 GTAACCAAGATATGTCAGTGTGACACTTGTGCGGGAGTGTCTATTGTGAAGCTGGC 2580
2672 aagaaacctgttaccatgaatcaatcaatgaagaacatcaatgataatgtaactgta 2731
2581 AAGAACCCGTTTACCATGAAATCATCATCTTGAACAATATCAGCTATGTGATAGCTA 2640
2732 tcaacgatgaatgataatgaagaagaatgtttggccagaacattcaaaagccctt 2791
2641 TCAACGATGAATGCTATCAAGTAAGACAAAGTGTTCACGAACCTTCAAAAGCCCTT 2700
2792 cccgttcaagcttccacttgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2851
2701 CCCGTTTACGCTTCCACTTGAATGATATGCAATCTGTCCCTTGTGCAAAAGATCTG 2760
2852 taagagaagaagatcactgtaagcaatgttggtaaaaaataatgaatgaatgaatga 2911
2761 TAAAGAGAGAAAGACTCATGCTAGGCAATGTTGGTGAATAATATGAATGAAGCCGG 2820
2912 agtatctgaagcagatgagctgttgaatgaagaatattgtctcttccacagaatg 2971
2821 AGTATCTGAAGCAGATGCACTGTTAGTGAATAATATGCTCTTCTTACCAAGATAG 2880
2972 ttgttccacatacaatctccttggcacatgacacagatattgtaagaatgacagata 3031
2881 TTGTTCCATATACAAATTCACCTTTTGGCACATGACCCAGATATGCAAAAGATGATA 2940
3032 ttgaacaacttaagaatgtaagaatgtcttgggttcttggtaaaataataatgagcta 3091
2941 TTGAACAACCTTAAAGATGTAAGTAAGTCTTGGTGTGGAATAATTAATGAAGCTGA 3000
3092 aaaaatgaataacagctacgcttcttcaagaagaatgtaagaatgtaagaatgtaaga 3151
3001 AAAATGAATAATACAGTCAAGCTTCTTATCAGAAAGATGTAAGTAATATTAACAAACA 3060
3152 aagaatgcccagaagacagaatgtaagaataatgaatgaatgaatgaatgaatgaatga 3211
3061 AAGATGCCCAAGACAGATGATGCAAAATGAATGAATAACTGTATACCTGTGTGATG 3120
3212 ttgcaatgaatcaatcactgtaagaatgtaagaatgtaagaatgtaagaatgtaagaat 3271
3121 TTGCCATGAAATATCATGATGCAAAAGATGATGATGATGATGATGATGATGATGATG 3180
3272 cagttactacagctgcttctcactcaactcaactcaactcaactcaactcaactcaact 3331
3181 CGGTACTACAGCTGCTTCTTCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 3240
3332 atctgctcctctgaatgaatcaatcttcaactcctggaataacccaataatgctc 3391
3241 ATCTGCTCTCTGAATAATGAATCATTTTTCATCTCGGAATAACCTTAACCAATGTTTC 3300
3392 taggaagctgttaacaagcacttcaatcaagaagcagaagcaatctcaagacaaatcaac 3451
3301 TAGGAGCTGTGTAACAAGCCATTTTATCAGAGGCAAGCAATCTCAGACCAATATATC 3360
3452 gaatgaagaatgtaagaatgtaagaatgtaagaatgtaagaatgtaagaatgtaagaat 3511
3361 GAATGGAATACTTAAGCAATGCAAGCAGAGCTCAAAATCAAGCTCTCTGGAAGATTA 3420
3512 aaggaagccttgaatgcttgaatgtaagaatgtaagaatgtaagaatgtaagaatgtaaga 3571
3421 AGGGAGGCTGATATGTTCTGAATGATCAGAGTAAGTAAGTAAGTAAGTAAGTAAGTCT 3480
3572 caacttgcggggaagaagaatgtaagaagaagaatgtaagaatgtaagaatgtaagaat 3631

REFERENCE 2 (bases 1 to 5110)
 AUTHORS Ohara, O., Nagase, T. and Kikuno, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-FEB-1999) to the DDBJ/EMBL/Genbank databases. Osamu Ohara, Kazuo DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail: odaiinfo@kazuza.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)

FEATURES

source

1. 5110
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="h07056"
 /clone_lib="pBluescriptII SK plus"
 /dev_stage="adult"
 /sex="male"
 /tissue_type="brain"
 1..4253
 /gene="KIAA0979"
 <1..4253
 /gene="KIAA0979"
 /product="KIAA0979 protein"
 /protein_id="BAA76823.1"
 /db_xref="GI:4589602"
 /translation="VRRLKVVVTFMDMDSDSEEEKLYLNLALHLSDFLKHPPKD
 VRLVACCLADIRIYAPAPYSPDKLIDFMETIROLKGLDMSPOBRYELLE
 NIAMWSYNICEFLEDSENEITOLYRYSVINGNHNOKWLMDSIIIGEDYV
 SOELDPTVNLVPAKHNKQAYDLAKALKTAQAIERYITNFRQVIMLKTSS
 DSEHVFDLLELYNIDSHLSLVLPQLEKLSNDNEELQVVKLAKNEKADSEL
 ASQNKPLMOCYGLRFNDIHVIRLECVKSHLMPLDLAKDLELKYRSDPEEA
 IRHDVLYITTAARDKLDLIVNDHLNFEVERTLDMRWYKEMAGLQYKYALQS
 AAGDAKQIAWIKMLKHLHYONSIDRLIVERIRVPHNLEETKMLTYLY
 ATDLIAVAKALNEMKKNLIRHOVLDLITKOPTDASVKRIFSKVMYITNLDNP
 GKADPMKFTQVLEDEDEKTRKQLEVVSCTCKQAGECVREITKLGKQPTNP
 LEMIKFLERIAVHIDTESISALIKQVNSIDGTADDEDEGVTQDAIRAGLELLY
 LSTPHRISPHSAITFESLACLKMDKEVAEALQIFKNTGSIEDFPRIKALLPV
 LHHKSKGPPQAKVAIHCHIAIFSSKEQFAQIFPLHSLDPSNLEHITPLVTG
 HIALPDQAPALKSIVATFIVKDLMDRLPGKTKTKLWPDDEVSPTMKVIAI
 KMWVRLLIGMKNHNSKSGTIRLLITLILSDGLDGGCISKPMDSRIPLASAIY
 KIAOEPCHYHITILEYOICALAINDECTIOVROYKRLKLSRLPLLEYAICAL
 CADPVERKRAHROCLVKNINRYRELYLOHAAVSKLSLLEEVVPTYLILLADP
 DYVKODIEQLADKQECLEFVLEILMAKNENSHARIRKAVENIKQTKDAQGDPAKM
 NEKLYVDCVAMNIIMSKSTYLSLESPKPLVAREFTQPKNFSNTKNLPLPEMKS
 FPGPKRTNNVGAIVNKLPLSAGKQSDTSSRMETVSNASSSSNPSPGKIKRDLT
 EMDHSENYETVSSPLPGKSKOKRSDSLVRSLEKPRGKTKPTVPEOEKRLGMDLT
 KLYOEOKPKGQSORSKRGRHASDEQOMPEERKRLDLENDDEONSPPKKRGRPR
 PPLPGGTPKEEFTKTSKSKGKKKSGPPAPEELEEEROSGTEOKSKOHRVSR
 AQQRAESPSSAIESTOSTPQKGRGRSPKTPSPSQPKKNVGRSKQAATKENDSSE
 VDFVFGSSPVDDIPQDETEEEESTVNVRRRSKRRRR"

BASE COUNT 1714 a 919 c 1045 g 1432 t
 ORIGIN

Query Match 97.6%; Score 4073; DB 85; Length 5110;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4076; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 92 tggtagagacataaagatggttgtagaaactttatgatgatagaccagagcctgaag 151
 |||||
 DB 1 TGGTGAACACATTAAAGATGTTGTGAACCTTTATGAGATATGACAGACACTCTTAAG 60
 QY 152 aagaaaagagccttattacacctaagccttacccttgcttgtagatctttttctcaag 211
 |||||
 DB 61 AAGAAAAGAGCTTTATTAACTTACCTTACCTTTCGCTTCAGATTTTTCGTAAGC 120
 QY 212 atcctggttaagatgctcgttaacgtagcgtcgtcgtcgtcgtatattcagagatt 271
 |||||
 DB 121 ATCCTGATAAGATGTCGCTTACTGTAGCCTGCTGCTTGCATATTTTCAGATTT 180
 QY 272 atgctctgaagctccttacaacatccctgataaataagagatatattatgtataa 331
 |||||
 DB 181 ATGCTCTGAAGCTCTTACACATCCCTGATAAATAAGAGATATATTATGTTATAA 240

QY 332 caagacagttgaagggttagagatacaaaagaccacaaatcaatgatttttatt 391
 |||||
 DB 241 CAACACAGTTCAAGGGCTAGAGATACAAAGACCCACAGTTCAATAGGATTTTATT 300
 QY 392 tacttgaagaacatgcttggttgaagtcacataaacaatgcttgtagaagataga 451
 |||||
 DB 301 TACTTGAGAACATGCTTGAGTCAAGTCATATATACATATGCTTGAGATTGAAATGCA 360
 QY 452 atgaaattttccaccacataacgaacatttttcaagttataaacaatgcccacatc 511
 |||||
 DB 361 ATGAAATTTTCACCACCTATACAGAACCTTATTTAGTTATAAACAATGGCCAAATC 420
 QY 512 agaaagtcacatgacatgtagaaccttataagctcattatttctggaagtcatacag 571
 |||||
 DB 421 AGAAGTCATATGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
 QY 572 tgcctcagagccttttgatacggctttagtaactcgtgtaacctgctcatalaagaattaa 631
 |||||
 DB 481 TGCTCAGAGACTTTTGATACGCTTGTAGTAATCTGTACCTGTATCAATGAAGATTTAA 540
 QY 632 acaagcagcatatgatttggaagccttctacgaagagagacgctcaagcattatgagc 691
 |||||
 DB 541 ACAAGCAAGCATATGATTTGGCAAGGCTTACTGAAGAGACGCTCAAGCTATTGAGC 600
 QY 692 cataattaccacattttttaatcaggttctgattcgtcttggaacaacatcacaagcatt 751
 |||||
 DB 601 CATATATTACCAATTTTATATAGCTTGCATAGCTTGGGAAACATCTATACGCCATT 660
 QY 752 tgcagaagcaltcttgacttaatttgagcgtcacaatattgataagcattgctgc 811
 |||||
 DB 661 TGTCAGAGCATGCTTGTACTTAATTTTGAGCGCTCAATATGATGATGATGATGATG 720
 QY 812 tctctgttttaacccacgtgaatttaataaagaagcacaatgaagagcgcctac 871
 |||||
 DB 721 TCCTGTTTAAACCCACCTGGAATTAATTAAGACCAATGATAAGAGAGGAGGCTTAC 780
 QY 872 aagttgttaaacctcgtgcaaaaatgcttggaaggaaggaattgagcttcctaaa 931
 |||||
 DB 781 AAGTTGTTAAACTCTGCGCAAAAATGTTTGGCGAAGGATTCAGATTTGGCTTCTAAA 840
 QY 932 acaagcactttgacgtgctacttggcaggttaataatgatalatccatccacatccgc 991
 |||||
 DB 841 ACAAGCAGCTTTGGCAGTGTACTTTGGGACAGTTTAATGATATCCATGATCCGCC 900
 QY 992 tggagatgtgaaatttgctgagccattgctcagaacacatccgatttggcaaaagact 1051
 |||||
 DB 901 TGGATGTGTAAATTTGCTAGCCATTGTCTCATGAACCAATCGATTTTACCAAAAGACT 960
 QY 1052 taacagagtaacttaaaagttagtcacatgacacgttggaagcatttagacatgattta 1111
 |||||
 DB 961 TAACAGAGTATCTTAAGTGTAGTGCACATGACCCGTGAGGAAGCATATTAACATATG 1020
 QY 1112 ttgtgcaatagttacagcgtcgttaaaagatalatcttctgctgcaatgatacattactta 1171
 |||||
 DB 1021 TTGTGTCATATGTTACAGCTGCTTAAAAAGATATCTTCTGTGTAATGATCATTACTTAA 1080
 QY 1172 atttgtagaagaagaacatttagaacaagaatggaagtagcgaagaagcagctatg 1231
 |||||
 DB 1081 ATTTGTGAGAGAGAGACATTAACAAACGATGAGAGTACGCAAGAAACCATGTATGG 1140
 QY 1232 gacttcccacaaatttaagaacatgcttataagcgaagcgtggaagaagatctgcaa 1291
 |||||
 DB 1141 GACTTGCACCAATTTATTAAGAAATATGCTTTACGTACGACGCTGGAAGATGCTGCAA 1200
 QY 1292 aacagatagcatggtatcaaaagacaatctgtacatalatalatalcaaaatagattatg 1351
 |||||
 DB 1201 AACGATAGCATGATGATCAAGACAAATTTGCTACATATATATTAACAATAATGATGATG 1260
 QY 1352 atgactacttgttgaaagcagatcttctcacaatcattggttcccaaatltgaaacta 1411
 |||||
 DB 1261 ATGCACTACTTTGTAAGCGGATCTTCTCATATCATGATGTTCCACAAATTTAGAAACTA 1320

QY 2341 ttgtctcctcttggaatcttgggtacttctatcttgtaagaatcttctcaagt 2400
 |||||
 Db 2406 ttgtgctgctctttgaaatctttgtaagctactttcttgtaanaatcttccatgaat 2465
 QY 2401 gatcgctccagggaaaaagaacaactaaacttgggtccagatgaagaatctct 2460
 |||||
 Db 2466 gatcgctccagggaaaaagaacaactaaacttgggtccagatgaagaatctct 2525
 QY 2461 gagacaatgtaaaaaatgaagctataaagtatgaatggttcggttacttgaaatga 2520
 |||||
 Db 2526 gagacaatgtaaaaaatgaagctataaagtatgaatggttcggttacttgaaatga 2585
 QY 2521 aataacacagtaaatcaggaacttctacattgaatgcttaacaacaatattgcaagt 2580
 |||||
 Db 2586 aataatcagtaaatcaggaacttctacattgaatgcttaacaacaatattgcaagt 2645
 QY 2581 gatgagacttgcaggaacaggggaaatagtaaacagataatgctacgtctgaact 2640
 |||||
 Db 2646 gatgagacttgcaggaacaggggaaatagtaaacagataatgctacgtctgaact 2705
 QY 2641 gctgctggagatgctattgtgaagctggaacagaaacctgttaccatgaatctacaa 2700
 |||||
 Db 2706 gctgctggagatgctattgtgaagctggaacagaaacctgttaccatgaatctacaa 2765
 QY 2701 ttagaacaatcagctatgctatgctatcgaacagatgctatcgaatgaaga 2760
 |||||
 Db 2766 ttagaacaatcagctatgctatgctatcgaacagatgctatcgaatgaaga 2825
 QY 2761 gctgttccagaaactcacaaggccttcccglttaacggtccacttgaatagat 2820
 |||||
 Db 2826 gctgttccagaaactcacaaggccttcccglttaacggtccacttgaatagat 2885
 QY 2821 ggaactgtgccccttctgcaaaagatcctgtlaaaggagaagaagctatgctaaagaa 2880
 |||||
 Db 2886 ggaactgtgccccttctgcaaaagatcctgtlaaaggagaagaagctatgctaaagaa 2945
 QY 2881 tgttctgtaaaaaataaataatgtaagcggaatctgaacagatgctatgct 2940
 |||||
 Db 2946 tgttctgtaaaaaataaataatgtaagcggaatctgaacagatgctatgct 3005
 QY 2941 gaaaaatcttctcttcttaccagatgctgttctccatatacaacttccactttggca 3000
 |||||
 Db 3006 gaaaaatcttctcttcttaccagatgctgttctccatatacaacttccactttggca 3065
 QY 3001 catgaccagatgctatgcaagatgcaagatgcaagatgcaagatgcaagatgca 3060
 |||||
 Db 3066 catgaccagatgctatgcaagatgcaagatgcaagatgcaagatgcaagatgca 3125
 QY 3061 ccttctgcttctggaatataatgcttgaagaatgaagaatgaagaatgaagaatgca 3120
 |||||
 Db 3126 ccttctgcttctggaatataatgcttgaagaatgaagaatgaagaatgaagaatgca 3185
 QY 3121 agaaagatgctgaagaataataaacaacaagaatgcccgaagaccagatgctcaaaa 3180
 |||||
 Db 3186 agaaagatgctgaagaataataaacaacaagaatgcccgaagaccagatgctcaaaa 3245
 QY 3181 atgaatgaagaactgctacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3240
 |||||
 Db 3246 atgaatgaagaactgctacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3305
 QY 3241 actacatacagcttgggaatcctcctaagaacccggtactacagcttcttctactcaa 3300
 |||||
 Db 3306 actacatacagcttgggaatcctcctaagaacccggtactacagcttcttctactcaa 3365
 QY 3301 cctggaagaatcttcaatgaaccccaaaaatattctgctccctggaatgaagaatcttctc 3360
 |||||
 Db 3366 cctggaagaatcttcaatgaaccccaaaaatattctgctccctggaatgaagaatcttctc 3425
 QY 3361 actcctggaagaactcaaaacaaacaaatgttctagaagctgtttaaagaacacttcatca 3420
 |||||
 Db 3426 actcctggaagaactcaaaacaaacaaatgttctagaagctgtttaaagaacacttcatca 3485

QY 3421 gcaagcaagcaatctcagaacccaatcatcagaatggaactgtlaagcaatgcaagcagc 3480
 |||||
 Db 3486 gcaagcaagcaatctcagaacccaatcatcagaatggaactgtlaagcaatgcaagcagc 3545
 QY 3481 agctcaatccaaagcttctcgtgaagaataaaggaggttgaatgttctgaatgtgat 3540
 |||||
 Db 3546 agctcaatccaaagcttctcgtgaagaataaaggaggttgaatgttctgaatgtgat 3605
 QY 3541 caagtgaaatgaagaatcacaatgttctacacttgcggggaaaaaagaatgacaag 3600
 |||||
 Db 3606 caagtgaaatgaagaatcacaatgttctacacttgcggggaaaaaagaatgacaag 3665
 QY 3601 agagaagactctgacttctgtaagctgtaattggaagccttagagcgagaaaaaag 3660
 |||||
 Db 3666 agagaagactctgacttctgtaagctgtaattggaagccttagagcgagaaaaaag 3725
 QY 3661 cccgtcacaagaacaggagaagaataggtatgagatgacttgaactgaatgttgaagaa 3720
 |||||
 Db 3726 cccgtcacaagaacaggagaagaataggtatgagatgacttgaactgaatgttgaagaa 3785
 QY 3721 cagaacacctaaagcgagctagcgaagctggaagaagcgcatagcgctcagaatctgat 3780
 |||||
 Db 3786 cagaacacctaaagcgagctagcgaagctggaagaagcgcatagcgctcagaatctgat 3845
 QY 3781 gacagcagctgagcttgaaggaagaagcgtcagaagaatata ttagaataatgaatga 3840
 |||||
 Db 3846 gacagcagctgagcttgaaggaagaagcgtcagaagaatata ttagaataatgaatga 3905
 QY 3841 cagaatagctccgcccgaataaagggttaaaaggcgcgaacccaacacttctgttgaggt 3900
 |||||
 Db 3906 cagaatagctccgcccgaataaagggttaaaaggcgcgaacccaacacttctgttgaggt 3965
 QY 3901 acaccaaaagagagcgaacaaatgaagaacttctaaagaagaagcaaaaaaatcttga 3960
 |||||
 Db 3966 acaccaaaagagagcgaacaaatgaagaacttctaaagaagaagcaaaaaaatcttga 4025
 QY 3961 cctcagaacagcaggaaggaagaagaagaagaacaagtgtgaatacggaaacagaag 4020
 |||||
 Db 4026 cctcagaacagcaggaaggaagaagaagaagaagaacaagtgtgaatacggaaacagaag 4085
 QY 4021 tccaaagaagaacagcaggaaggttcaaggaagaagcagaagcagaatctcttga 4080
 |||||
 Db 4086 tccaaagaagaacagcaggaaggttcaaggaagaagcagaagcagaatctcttga 4145
 QY 4081 tctagtgcaatggaatccacacagctccacacagaagaagcaggaagacatcaaaa 4140
 |||||
 Db 4146 tctagtgcaatggaatccacacagctccacacagaagaagcaggaagacatcaaaa 4205
 QY 4141 acgcatcacatcacacaacaaaaaaatgtg 4173
 |||||
 Db 4206 acgcatcacatcacacaacaaaaaaatgtg 4238
 |||||
 RESULT 3
 AB023196 5110 bp mRNA PRI 16-JUN-1999
 LOCUS Homo sapiens
 DEFINITION Homo sapiens mRNA for KIAA0979 protein, partial cds.
 ACCESSION AB023196
 VERSION AB023196.1 GI:4589601
 KEYWORDS
 SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_lhb:pb1uescript11
 SK plus clone:hj07056.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (sites) Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
 Miyaajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
 Prediction of the coding sequences of unidentified human genes.
 XIII. The complete sequences of 100 new cDNA clones from brain
 which code for large proteins in vitro
 JOURNAL DNA Res. 6 (1), 63-70 (1999)
 MEDLINE 99246063

186 ACCTTTATGATATGAGACGAGACTGTGAAGAAAGAGGCTTTATTTAAACCTAGCT 245
181 ttacattcttcaagaattttttctcaagaatcttgtaaagaatttcgcttactgta 240
246 TTACATCTTGCTGATGATTTTCTCAAGCATCTCTGATTAAGATGTTCGCTTACTGTA 305
241 gctctgcctctgctgataatttcaagattatgctctcgaagctccttacaacccct 300
306 GCCGTGCGCTTGCTGATTTTCAAGATTTATGCTCTGAGAGCTCTTACACATCCCT 365
301 gataactaagaatataattatgatttataaagaagcttgaagggtctagagataca 360
366 GATTAATCAAGATATATTTATTTATTAACAAGACAGTGAAGGGGCTAGAGATACA 425
361 aagaagcccaaatcaatagatttttatttacttgaagaacttggttgtaagtca 420
426 AAGAGCCCAAGCTCAATAGATATTTATTTGCTTGAAGACATTTGCTGGGTCAGTCA 485
421 tatacatacttctgtagtagaagatagaatgaatlttaccacgctatagaagacc 480
486 TATACATATGCTTTGATTTAGATTAAGATAGCAATGAATTTTACCCAGCTTACAGAAC 545
481 ttatttcaagttataaacaatgcccacaatcagaagtcacatgacatgtagacctt 540
546 TTATTTTCAAGTTATTAACAATGGCCACATCGAAGTCATATGCAATGATGAGACTT 605
541 atgagctctatatttctgaaagtgatagacagtgctcgaagcttcttgaaagcttta 600
606 AAGAGCTCTATTTATTTGGAAGGTGATACAGTGTCTCAGAGCTTTTGATACGGTTTAA 665
601 gtaaatctgtacacctgctcacaagaatttaacaagaagaatataatgctcgaagct 660
666 GTAATCTGGTACTGCTGCATTAAGATTTAAACAAGACATATGATTTGGCAAGGCT 725
661 ttactgaagaagacagctcaagctatgagccatataattaccactttttaaacaagtt 720
726 TTACTGAGAGGACAGCTCAACGCTATTGAGCCATATTTTACCAATTTTAAATCAGGTT 785
721 ctgagcttgtagaaacaatcacaagcatttgcagaagcagtgcttgtaatttttg 780
786 CTGATGCTTGGAAGAACCTCATACGCAATTTGTCAGAGCATGCTCTGACTTTAATTTTG 845
781 gagctcacaatatgtagtcatcttgctcctcgtcttaccacagctgaaatttaa 840
846 GAGCTACATATTTGATAGTATGATGCTGCTGCTGCTTTTACCCAGCTGCAATTTAA 905
841 ttaagaagcaatgaataatgagagcgctcacaagttglttaaactactgycgaataatgctt 900
906 TTAAGAGCAATGATTAAGAGAGCGCTTACAGTGTTAACATACGCGCAAAATGTTT 965
901 ggggcaagaagattcagaattgctctcacaagaagccacttggcagtgtaattgagc 960
966 GGGCAAGGATTTAGATTTGCTTCTCAAAACAGCCACTTGTGCACTGCTACTGGGC 1025
961 aggttaattatatacctatgtaaccacccgctcgtgaatgltgtaaatctgtaagcattgt 1020
1026 AGGTTATGATATTCATGATGATGCAATCGCTGGAATGTGGAATTTGCTAGGCATTTGT 1085
1021 ctaatgaacacatcctgatttagcaaaagacttaacagagatctctaaagtgaagtcacat 1080
1086 CTCATGAGACATCTCTGATTTTGAAGAAAGACTTAACAGAGTATCTTAAAGTGAAGCTACAT 1145
1081 gaccctagaagaagctatagaagatgattatgttcaaatgaattacagcgtcctaaagaa 1140
1146 GACCTTAGAGAGCTATTTAGCATGATGATTTATTTGCTAAATAGTTACAGCTGCTTAAAG 1205
1141 gatacttctctgcaatgtagtcaacttaacttaattctgtaagaagagagataatgacaaa 1200
1206 GATATTTCTTCTGCTCANTGATGATCTTACTTATTTTGGGAGAGAGAACATTTAGACAAA 1265
1201 cgaatgagagtagcaagaagaacatgtaggagcttgcccaaatltaagaatatgct 1260
1266 CGATGAGAGTAGCCAAAAGACCATGATGGAGCTTCCCAAAATTTATTAAGAAATATGCT 1325
1261 ttacagtcagcagcttggaataaagatgctgcataaacaagatagcatgtagtcaagaacaa 1320
1326 TTACAGTCAGCAGCTGGAAGAAAGATGCTGCAAAACAGATAGCATGATCAAAAGCAAAATTG 1385
1321 ctacatatatactcaaaatagatattgtagtgcagacttctgtaagcagatcttgc 1380
1386 CTACATATATATTTCAAAATAGATTTGATGATGATGATGATGATGATGATGATGATGAT 1445
1381 caatacatggttctctcaaatattagaactcagaagagatgaatgtaattatattcttg 1440
1446 CAATACATGTTGCTCTCAAAATTTAGAACTCAACAGGAGGAAATCTTATATTACTTG 1505
1441 tatgcacactggaatttaaatgctgttaaaagcatltagatgaatgtaattgtaattgcaaat 1500
1506 TATGCCACACTGATTTTAAATGCTGTGAACCATTTGATGAATGAATGTAATGTCANAT 1565
1501 ctgctccgacalcaagaataaagatttctgacttgaattaaagcaaccacaagatgct 1560
1566 CTGCTCCGACATCAAGTAAAGATTTGCTGATTTGATTAAGCAACCCAAACAGATGCC 1625
1561 agtctcaagcacaatttcaaaaagtatggttattacaagaatttaactctgctgct 1620
1626 AGTGTCAAGGCCATATTTTCAAAAGTGTGTTTACAAAGATTTTAACTGATCTGCT 1685
1621 aagctcagaattcacaagaataatcacacaggttattagaagaatgtagaagaataaga 1680
1686 AAGCTCAGGATTTTCAATGAGAAATTTCAACAGAGTGTAGACAGTATGATGACAAATTAAGA 1745
1681 aagcagttagaagacttctgtagtccacaatctgctcctgcaagcagctgaaagttgtg 1740
1746 AAGCAGTTAGAAAGTACTTGTATGTCACAAATGCTCTGTCAGACAGGCTGAAGGTTGTG 1805
1741 cgtgaataaactaagaagttggtgcaaccacaagaagctcaaaatctcttctgtaagt 1800
1806 CGTAAATTAACATTAAGATTTGGCAACCCCAAGAGCTTCAAAATCTTCTTCTGGAATG 1865
1801 atcaagttctcttgtaggaagatagacacctgtagacataagatccagatcactagct 1860
1866 ATCAAGTTTCTTGTGAGAGAGATAGCACTGTGCACATAGTATCCGATCTATATAGTCT 1925
1861 ctatlaaacaagtgaaacaatcaatagatgaaacagcagaatgaatgaagtgagtgct 1920
1926 CTTATTAACAAAGTGAACAAATCATATGATGAGAACAGACATGATGAAGATGAGGATG 1985
1921 ccaactgatacagccatcagaagcagcttgaactgcttgaagttactctcattacaat 1980
1986 CCAACTGATCAAGGCATCAGAGAGGCTTGAACCTGTTAAGGTATCTCATTTTACACAT 2045
1981 ccaatcactatcactcgtcgaagaacatttgaatcattactggtgctgctgaagaatgag 2040
2046 CCATCTCATTTTCAATCTGCTGGAAGCAATTTGAATCATTTACTGCGCTGCTGTAATAATGAT 2105
2041 gatgaaaaagtagaagaagctgcaatcaaaatttcaaaaacacagaagaagaattgaa 2100
2106 GATGAAGAAAGTAGCAGAAAGCTGCACTACAAATTTTCAAAACACAGGAAGCAAAATTGAA 2165
2101 gagatattcacaacatcaagatcagctgctgctgctgctgcttcaacaataatcaaaaaa 2160
2166 GAGATTTTCCACACATCAGATCAGCTTGTCTCTGTTTATCATCAAAATCTTAAAAAA 2225
2161 ggaaccccccgctcaagaacaaatgcaatctgaattgtaacatgagataatttctagtaa 2220
2226 GGACCCCCCGCTCAAGCAAAATATGCTTATTTGATATGATGATGATGATGATGATGATGAT 2285
2221 gagaacccagtttgcacagataattgagcctctgtaagaagcctagatcccaagaacctg 2280
2286 GAGGCGCAGTTTGGACAGATATTTGAGGCTCTGATTAAGACCTAGATCCAAAGCAACCTG 2345
2281 gaacatctataacaacatggttactatggtcatatatgtctcctctgcaactgataa 2340
2346 GAACATCTCATTAACACATTTGTTACTATTTGTCATATTTGCTCTTCTGCACTGATCAA 2405

CDS

/number=2
66..4241
/gene="AS3"
/codon_start=1
/product="androgen-induced prostate proliferative shutoff
associated protein"
/protein_id="AND2134.2"
/db_xref="GI:4559410"
/translation="MAHSTRTNDGKITPPGCKEISDKISKENAVRLKMYKTFMD
MDQSEEEKELYLNALHLSDFLKHDPKDVRLVACCLADIFRIYAPVPEAYTSPDK
LKDIEMFTIRLOLKGLEDTRKSPQFNRYFYLENLIAWKGVNICFELEDNEIETOLYRT
LEFVYINNGHNGVHMWDLMSSITCEDTYSQELLDVYVNLVPAHKNLKOATDILA
KALKRTAOATEPIITNFNOVLMGKTSISDLSHVSDDLIELYNIDSHLSVLPO
LEFKLSNDNEERLQVVKRLAKMFGAKDSELAQKPLMQCYLGRFNDIHPVIRLECY
KFASHCLNNHDLADLLEYLKVRSHPDEEALRHVIVISYTAAKDILLVNDILNF
VGEYLRDKRMVRYKEMAMGLAOIYKKVALQSAAGDAKQAIWIKDKLHITYONSID
DLIYERLIFAOYVPHNLETTERMKCLYIYATLDLNAVKALNEMWKKCONLIRYQDIL
LLDLTKQKTDASVKAISKVYVITRSLPDRKADPPMKKTQVLEDEDKIRKQLEVL
VSPTCCKOAEQVREITRKUGNKPQPTNPLEMKFLLERIAPVHIDESTISALIKO
VNSISDGTADDEGVPTDOAIRAGLELLKYLSTFHPISFHSAEFFESILACLKMDDE
KVAEALQIFPKNTGSKIEDFPHRSALPYLHHSKRGPPROAKYAIHCIIAIFSSK
EAOFOAIEPLPKSLDPSNLEHLITPLVTIGIALLADDPAPLKSIVATPIYDIL
MNDRLPGKTKTKIWPDEEVSPETPVKIOAIKMYRMLLGMKNHSGSTRLTLTT
ILHSDGLTEBGKISKPDMSRLRLAAGSAIVKLAQPCYHEITILEYQOLCLAIINDE
CYQVQVFAOKLHKLGLSLRLLPLEYMAICACAPVERERHARQCLVKNINVRREY
LKQHAASVEKLISLLEPYVPTIHLAHADPYVYODIEQLKDYKECLWFLIELIMA
KNENSHAFIRKMYENIKOTKDAQDPDAKMEKLYTCDVAMNIIMSKSTYSLESP
KDPVLPAREFTQDPKFNFSNTKNYLPKPGKSFETPGKPTTNVLGAVNKLPSAGKQSO
TKSSRMETVGNASSSSNPGRIGKGRDSSPMDSHENDYTSSPLGKRSKDRDSS
DLYRELEKPRGRKTPYTEREKEKGMIDLTKLVQEQRPKGSQSRKRGHNASSEDO
QMPERKRLKEDLLENEDONSPPKGGKGRPKPLGGSTPEEPTMKSKGSKRSG
PPAPEEEEEESGNTQKSKSKOHVRSRAQRAESESSESAIESTOSTPOKGRNP
SKTSPSQPKKNV"
174..377
/gene="AS3"
/number=3
378..464
/gene="AS3"
/number=4
465..562
/gene="AS3"
/number=5
563..689
/gene="AS3"
/number=6
690..770
/gene="AS3"
/number=7
771..911
/gene="AS3"
/number=8
912..1027
/gene="AS3"
/number=9
1028..1122
/gene="AS3"
/number=10
1123..1268
/gene="AS3"
/number=11
1269..1420
/gene="AS3"
/number=12
1421..1534
/gene="AS3"
/number=13
1535..1616
/gene="AS3"
/number=14
1617..1665
/gene="AS3"
/number=15
1666..1805

/gene="AS3"
/number=16
1806..1921
/gene="AS3"
/number=17
1922..2027
/gene="AS3"
/number=18
2028..2188
/gene="AS3"
/number=19
2189..2312
/gene="AS3"
/number=20
2313..2471
/gene="AS3"
/number=21
2472..2540
/gene="AS3"
/number=22
2541..2677
/gene="AS3"
/number=23
2678..2801
/gene="AS3"
/number=24
2802..3006
/gene="AS3"
/number=25
3007..3121
/gene="AS3"
/number=26
3122..3254
/gene="AS3"
/number=27
3255..3374
/gene="AS3"
/number=28
3375..3437
/gene="AS3"
/number=29
3438..3583
/gene="AS3"
/number=30
3584..3689
/gene="AS3"
/number=31
3690..4129
/gene="AS3"
/number=32
4130..4354
/gene="AS3"
/number=33
4355..4810
/gene="AS3"
/number=34

BASE COUNT 1755 a 944 c 1074 g 1480 t
ORIGIN

Query Match 99.2% Score 4141; DB 97; Length 5253;
Best Local Similarity 99.5% Pred. No. 0;
Matches 4153; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 atggctcattcaaaactagagccaatgataagaaaattacatccgctggggtcaag 60
|||||
DB 66 ATGGCTCATTCAAAGACTAGACCAATGATGGAATAATTACTATCCGCTGGGTCAAG 125
|||||

QY 61 gaatatcagataaaatctaaagagagatggtgagacgataaagattgtgaa 120
|||||
DB 126 GAAATATCAGATAAATATCTAAAGAGAGATGTCGAGACGATTAAAGTGGTTGTGAAA 185
|||||

QY 121 accttattgataltgaccagagactctgaagaagaagaagagacttatttaaacctagct 180

Db 2947 GCATCTGTCCTTGTGCAAAAAGATCTGTAAGAGAGAAAGAGCTCATGTCAGCAAA 3006
 2881 ttttttgtaaaaaataataatgttaagcggagatcttgaaagcagctgcagcgttagt 2940
 Db 3007 TGTGTGTGAAAAATATTAATATGAAGCGGAGATCTGTAAGCAGCATGCGCTTTAGT 3066
 2941 gaaaatatatgtctctctccagagatgtgttcataatacattacccttttgca 3000
 Db 3067 GAAAAATATTATTTGCTCTTCAACGAGATGTTTCCATATACAAATTCACCTTTTGCA 3126
 3001 catgaccagattatgtcaaatgaagtaagatattgaacaacttaagaatgtttaaagt 3060
 Db 3127 CATGACCCAGATATGTCAAAAGTACAGATATGAAACAATTAAGATGTTAAAGATGT 3186
 3061 cttgtgtttgtctgtaaatattatgtcttaaaaaatgaataacagctcaagctttatc 3120
 Db 3187 CTTTGGTTTGTCTGGAATAATTAATGCTAAAAATGAAATTAACAGTCAAGCTTTTATC 3246
 3121 agaaagtgtgtaaaaatatatacaacaagaatgccagagaccagatgtgcaaaa 3180
 Db 3247 AGAAAGATGTAAGAAATATTAAACAAACAAAGATGCCAAGAGACGATGATGCAAAA 3306
 3181 atgaatgaaaaaactgtacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3240
 Db 3307 ATGAATGAAAAAATGTACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3366
 3241 actacatacagttgtggaatctccctaaagaccggtactaccagctgttttcttcaaca 3300
 Db 3367 ACTATCAATACAGTTTGGATCTCTTAAGACCCGGTACTACAGTCTGTTTCTTCACTCA 3426
 3301 cctacagaattcagtaacacacaaaatatctgcctccctgtaaatgaatacttttc 3360
 Db 3427 CCTACACAGAAATTCAGTACACCAAAAATTAATCTGCTCTGTAATTAATTCATTTTTC 3486
 3361 actcctgtgaaaaacctaacaacacaaatgttcttagagctgtttaaagccacttcalca 3420
 Db 3487 ACTCTGGAAGAACTTAACCAACCAATGTTCTAGAGCTGTTAACAAAGCCCTTTATCA 3546
 3421 gcagagcagaatctctgacccaataatcaccgaatgtgaatgaagaatgcaagcagc 3480
 Db 3547 GCAGGCAAGCAATCTGACACCAATCATCAGAAATGGAATGGAAGCAAGCAAGC 3606
 3481 agctcaaatccaagctctcctgtgagaataaaggagagctgtgtagtctgtaaatgat 3540
 Db 3607 AGCTCAATTCACACCTCTCTCGGAGAAATTAAGGAGGAGCTTGATAGTTCTGAATGAT 3666
 3541 cacegtgaaaaatgaagattacacaaatgtctcacccttgcgggggaaaaaagtgcacag 3600
 Db 3667 CACAGTGAATAATGAAGATTACACAAATGCTTCACCTTGGCGGGAAAAAAGTGACAAAG 3726
 3601 agagaagactctgtcttctgttaaggtctgaatgtgagaagccttaagaggaagaaaaacg 3660
 Db 3727 AGAAGACACTCTGTGATCTTGAAGGCTGGAATTTGAGAGGCTTAAAGGAGGAAAAAAGC 3786
 3661 cccgtcacagaagagaggaagaaattaggtatgtgactgagctgaagtgtgtgacagaa 3720
 Db 3787 CCGCTCACAGAACAGAGAGAAATTAAGTATGATGATGATGATGATGATGATGATGATGAT 3846
 3721 cagaacctaaagcagctcagcgaagtcgnaaaagagccataagcttcagaaatctgat 3780
 Db 3847 CAGAAACCTTAAGGCACTACAGAAATGCGAAAGAGGCCATTACGGCTTCAGAAATCTGAT 3906
 3781 gaaagcaggt 3840
 Db 3907 GAAAGAGAGTGTGCTGTGAGGAAAGAGGCTCAAGAGATATATTAAGAAATGAAGATGAA 3966
 3841 cagaatgtccgcaaaaaaaggttaaaagagcgcaaccacaaacccctgtgtgtgtgtgtgt 3900
 Db 3967 CAGAAATAGTCCGCCAAAAAGGTAAGAGGCGGACCAAAACCTCTTGCTGTGAGGT 4026
 3901 acacaaagaagagccacaaatgaaaacttaaaaaaggaagcaaaaaaaatctgga 3960

Db 4027 ACACCAAGAGAGCCACATGATAAACTCTTAAGAAAGCAACCAAAAAATCTGGA 4086
 3961 cctcagcaccagagagagaggaagaagaagaagaagatgtgaaataccgaacagaag 4020
 Db 4087 CTTCCAGCACAG 4146
 4021 tccaaagcaaacagcagcagagtgtaagagagcagacagagagagagatctctgaa 4080
 Db 4147 TCCAAAGCAAGACAGACCGAGTGTCAAGAGAGACACAGCAGAGAGAGAGATCTCTGAA 4206
 4081 tctagtcaattgaatccacacagtcacacacacagaagaagaagagagacatcaaaa 4140
 Db 4207 TCTAGTCAATTTGAATCCACAGCTCCACACCCAGAAAGAGAGAGAGAGAGAGAGAGAG 4266
 4141 agccatcacatcac 4172
 Db 4267 AGCCATCACCATCACAACCAAAAAATGT 4298

RESULT 2
 LOCUS HSU95825 5253 bp mRNA PRI 05-APR-1999
 DEFINITION Human androgen-induced prostate proliferative shutoff associated protein (As3) mRNA, complete cds.
 ACCESSION U95825
 VERSION U95825.2 GI:4559409
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 5253)
 REFERENCE
 AUTHORS Geck, P., Szelei, J., Jimenez, J., Soto, A.M., and Sonnenschein, C.
 TITLE Androgen-induced proliferative shutoff in prostate cancer cells
 JOURNAL Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
 2 (bases 1 to 5253)
 REFERENCE
 AUTHORS Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C., and Soto, A.M.
 TITLE Early gene expression during androgen-induced inhibition of proliferation of prostate cancer cells: a new suppressor candidate on chromosome 13, in the BRCA2-Rb1 locus
 JOURNAL J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
 3 (bases 1 to 5253)
 REFERENCE
 MEDLINE 99229875
 AUTHORS Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C., and Soto, A.M.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
 4 (bases 1 to 5253)
 REFERENCE
 AUTHORS Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C., and Soto, A.M.
 TITLE Direct Submission
 JOURNAL Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
 REMARK
 COMMENT Sequence update by submitter
 On Apr 5, 1999 this sequence version replaced gi:4539617.
 FEATURES
 location/Qualifiers
 1..5253
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="13q12-q13"
 /cell_line="LNCaP"
 /cell_type="cancer cells"
 /tissue_type="prostate"
 /note="exon 1 maps to P1 Artificial Chromosome PAC26H23; exons 2-24 map to cosmid 267p19; exons 25-34 map to P1 Artificial Chromosome PAC 49j10"
 1..5253
 /gene="AS3"
 1..46
 /gene="AS3"
 /number=1
 47..173
 /gene="AS3"
 exon

| | | | |
|----|------|--|------|
| QY | 1741 | cgjgaataactaagaagtgtgggcaaccccaacagcctacaatcccttcctggaatg | 1800 |
| Db | 1867 | CGTGAATAAATAAGAAAGTTGGGCAACCCCAACACCCTACAAATCCTTTCCTGGAAATG | 1928 |
| QY | 1801 | atcaagtttcctctggagagatagacccctgtgcataatagataccgaatctcaagtgct | 1860 |
| Db | 1927 | ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTCT | 1988 |
| QY | 1861 | cttatcaacaagtgaacaaatcaatagatlgaaacagcagatgatagaatgagggtgtc | 1920 |
| Db | 1987 | CTTATTAAACAAAGTGAACAAATCAATAGATGAGAACAGCAGATGATGCAAGTACGGGTGT | 2046 |
| QY | 1921 | ccaactatcaagccatcaagaagagagctcttgaactgtcttaagagtaactctcattacaat | 1980 |
| Db | 2047 | CCAACTGATCAAGCATACAGCAAGCGCTTGAACTCTTAAGTACTCTCATTTACACAT | 2106 |
| QY | 1981 | cccatcctcatctcactctgtcttgaacaaattgaaactaacttactgtgtcttgaanaatgtat | 2040 |
| Db | 2107 | CCCATCTCATTTCTCACTCTGCTGGAACATTTGAATCATTTACTGGCTTGTCTGAAATGTAT | 2166 |
| QY | 2041 | gatgaanaagtaagaagaagctgtgcatacaaaatttcaaaaaacagggaagcaaaattgaa | 2100 |
| Db | 2167 | GATGAAAAAGTAGAGAAAGCTGCATCACAAAATTTTAAAAACACAGGAACAAAATTTGAA | 2228 |
| QY | 2101 | gaggaatttccacacacatcagatcagccctgtctctctgttttcaatcaacaactcaaaaaa | 2160 |
| Db | 2227 | GAGGATTTTCCACACATACACATCAGCCTTGCTCCCTGTTTACATCACAAATCTAAAAA | 2286 |
| QY | 2161 | ggaeccecccgctcaagcnaatatgccaatcatgtlatccatgtcgatatlttctagtaa | 2220 |
| Db | 2287 | GGACCCCCCTCTCAAGCAAAATATGCGCATTTCAATGTATCCATACGCGATATTTTCTGTATAA | 2346 |
| QY | 2221 | gagagccagatttgaacagatatlttgagccctctgcataaagagcctagatcccaagcaacctg | 2280 |
| Db | 2347 | GAGACCCAGTTTGCACAGATATTGAGCCCTCTCATTAAGCCCTATATCCAAAGCAACCTG | 2406 |
| QY | 2281 | gaacatctcaataacacatctgttactatgtgtcatatgtctctcccttgcacctgataca | 2340 |
| Db | 2407 | GAACATCTCATTAACACACCATTTGTTACTAATGTGCATATGTGCTCTCTCCATCAGCTATCAA | 2466 |
| QY | 2341 | tttgctgtcctcttggaaatcttgggtagcacttcatgttgaagaatctctctctatgat | 2400 |
| Db | 2467 | TTTCTGTCTCTTGGAAATCTTTGGTAGGCTACTTTCATTTGTAAAGATCTTCTCATAGAT | 2528 |
| QY | 2401 | gacgcgtctccagggaaaagaacaaacttaaaccttgggtctccagatgaaagaatctatcctc | 2460 |
| Db | 2527 | GATGGGCTTCCAGGAAAAAGACACATTAACCTTTGGGTTCCAGATGAAAGATATCTCTCT | 2586 |
| QY | 2461 | gagacaatgtgtcacaatctcaggtctatataaaygaatgtgtcgatgtgctacttggaaatgaa | 2520 |
| Db | 2587 | GAGCAATAGTCAAAATTTAGCGCTATTAAATATGTTGTCGATGGCTTATGTGAATGAAA | 2646 |
| QY | 2521 | aataatcaagtaaatcaagaactcttacccttaagaatgtctcaacaacaatatgtgaatgt | 2580 |
| Db | 2647 | AATATATCACAGTAAATATAGGAACCTTCTACTTAAGATTGCTTAACAACAATATTCGATAGT | 2706 |
| QY | 2581 | gaatgagacttgcacagaacagaygggaaatagttaaacagagatagtltcaagtlctgagactt | 2640 |
| Db | 2707 | GATGAGAGCTTGAACAGAAAGGAGGAAAAATTAGTAACACAGATATGTCAAGCTGTGAGACTT | 2766 |
| QY | 2641 | gctgcctggagtgctatgttgaagctctggacaagaacccctgttaccatgaaatctacaa | 2700 |
| Db | 2767 | GCTCTGGAGTGTCTATTGTGAAGCTGTGGCAAGAACCTGTTCACATGAATAATCTATCA | 2826 |
| QY | 2701 | ttagaacaatatcaagcctatgtcatagctatatacaagaatgaaatgtcatatcaagtaagaaca | 2760 |
| Db | 2827 | TTAGACAATATCATGACTATGTGCAATTAGCTATCAACAGATGAATGCTATCAATAAGACAA | 2886 |
| QY | 2761 | gtgtgttcccaagaactctcaaaagccttcccggtttagcggctcccaactgtagatagt | 2820 |
| Db | 2887 | GTGTGTTCGCCAGAAACTTTCACAAAGGCTTTTCCCGTTTACGGGCTTTCACCTTGAGATATVG | 2946 |
| QY | 2821 | gcaatctgtgccccttltgtcaaaaagatccctgtataaagaagaagaagctcatagtctaagcaa | 2880 |

exon /number=20
2374..2532
/note="matches EST D20082 from clone pm1283."
exon /number=21
2533..2601
/note="matches EST AW212954 from clone IMAGE:2647801"
exon /number=22
2602..2738
/note="matches EST A177347 from clone 621810"
exon /number=23
2739..2862
/note="matches EST AW048929 from clone
UI-M-BH1-amo-c-09-0-UI"
exon /number=24
2863..3067
/note="matches EST A154932 from clone 599194"
misc_feature /number=25
/complement(2912..3302)
/note="matches EST AA630616 from clone 856164"
exon /number=26
3068..3182
/number=26
3183..3315
/number=27
3302..3572
/note="matches EST AA333118"
exon /number=28
3349..3818
/note="matches EST AA282190 from clone IMAGE:704631"
misc_feature /number=29
3436..3498
/number=29
3499..3644
/number=30
/complement(3586..4065)
/note="matches EST A1655429 from clone IMAGE:2239331"
exon /number=31
3645..3750
/number=32
/complement(3864..4286)
/note="matches EST A1680124 from clone IMAGE:2264521"
misc_feature /number=33
4001..4119
/note="matches EST N83765 from clone KK2745"
exon /number=34
4135..4450
/note="matches EST F12354 from clone c-38h11."
misc_feature /number=35
4191..4298
/number=35
4255..4307
/note="matches EST T64867 from clone 22264"
exon /number=36
4299..4434
/number=36
/join(4398..4533,4518..4660,4659..4675,4674..4693)
/note="matches EST H44781 from clone 188032"
misc_feature /number=37
4435..7444
/number=37
4537..5022
/note="matches EST A1681422 from clone IMAGE:2272632"
exon /number=38
4537..4892
/note="matches EST AA911125 from clone IMAGE:1416662"
misc_feature /number=39
4538..4969
/note="matches EST A1275866 from clone IMAGE:1877482"
exon /number=40
4538..4947
/note="matches EST AA987361 from clone IMAGE:160350"
misc_feature /number=41
4538..4943
/note="matches EST A1299646 from clone IMAGE:1898010"
exon /number=42
4562..5003
/note="matches EST AA447433 from clone 784555"
misc_feature /number=43
4571..4854
/note="matches EST H93424 from clone 220700"
exon /number=44
/join(4573..4649,4650..4815)
/note="matches EST W20070 from clone 306049"
misc_feature /number=45
/complement(4529..4937)
/note="matches EST AW089775 from clone IMAGE:2593146"
exon /number=46
4746..4994
/note="matches EST AA747568 from clone IMAGE:1269696"
misc_feature /number=47
/complement(4803..5332)
/note="matches EST A1911784 from clone IMAGE:2329873"
misc_feature /number=48
/complement(4828..5296)
/note="matches EST AA563884 from clone IMAGE:1013990"

misc_feature 4861..5243
/note="matches EST D20082 from clone pm1283."
misc_feature /complement(join(4872..4921,5281..5332))
/note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature join(4872..4921,5281..5332)
/note="matches EST A177347 from clone 621810"
misc_feature /complement(4872..4921)
/note="matches EST AW048929 from clone
UI-M-BH1-amo-c-09-0-UI"
misc_feature 4872..4921
/note="matches EST A154932 from clone 599194"
misc_feature /complement(4876..5333)
/note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature /complement(join(4922..4975,5281..5332))
/note="matches EST A1012649 from clone RPLBA02"
misc_feature /complement(4922..4975)
/note="matches EST AA899586 from clone
UI-R-E0-cy-e-02-0-UI; matches EST A1172172 from clone
RM0BV82"
misc_feature /complement(4931..5332)
/note="matches EST AA625961 from clone 745478"
misc_feature /complement(4951..5332)
/note="matches EST A1140790 from clone IMAGE:1566222"
misc_feature /complement(5021..5309)
/note="matches EST H45693 from clone 188292"

Query Match 99.8%; Score 4165.6; DB 92; Length 7444;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atggctcattcaaaagactagaccatgatgaaataatcatalccgcctggtggtcaag 60
127 ATGGCTCATTTCAAGACTAGACCATGATGAAATAATTCATATCCGCTGGGCTCAAG 186
Db 127 ATGGCTCATTTCAAGACTAGACCATGATGAAATAATTCATATCCGCTGGGCTCAAG 186
QY 61 gaataatcagaataaataatcagaagaagatggtgtagacgataagaatggtgtaaa 120
187 GAAATATCAGATTAATAATATCTAAAGAGAGATGCTGAGACATTAAAGATGCTGTGAAA 246
Db 187 GAAATATCAGATTAATAATATCTAAAGAGAGATGCTGAGACATTAAAGATGCTGTGAAA 246
QY 121 actttatgtagatgagccagactctgaaagaagaagagacttttaaacctagct 180
247 ACTTTATGATATGAGCCAGACTCTGAAGAAAGAGGCTTTATTAAACCTAGCT 306
Db 247 ACTTTATGATATGAGCCAGACTCTGAAGAAAGAGGCTTTATTAAACCTAGCT 306
QY 181 ttacatctgcttcagattttttcctaagacatcctgtaagaatgctcgtactgta 240
307 TTACATCTGCTTCAGATTTTCTCAGACATCTGTAAAGATGTTCCGTCTAGTGA 366
Db 307 TTACATCTGCTTCAGATTTTCTCAGACATCTGTAAAGATGTTCCGTCTAGTGA 366
QY 241 gectgctgcttgcgtatattcagaatltatgctcctgaagctccttacaacccct 300
367 GCTGCTGCTTGCCTGATATTTTACAGATTTATCTCCTGAAGCTCCTTACACATCCCT 426
Db 367 GCTGCTGCTTGCCTGATATTTTACAGATTTATCTCCTGAAGCTCCTTACACATCCCT 426
QY 301 gataaactaaagatatattatlttataacaagacagttgaagggctagaaggtata 360
427 GATTAACCTAAAGATATATTTATGTTTATATACAGACAGTTGAAGGGCTAGAGATTA 486
Db 427 GATTAACCTAAAGATATATTTATGTTTATATACAGACAGTTGAAGGGCTAGAGATTA 486
QY 361 aagagccacaatcaatagatlttatttacttgaagaacatgcttgggtcaagltca 420
487 AAGAGCCACAATCAATAGATTTTATTTACTTGGAACATTCGCTGGGCTCAAGTCA 546
Db 487 AAGAGCCACAATCAATAGATTTTATTTACTTGGAACATTCGCTGGGCTCAAGTCA 546
QY 421 tataacataatgcttgaagtagaagaatgaatattccaccagctataagaacc 480
547 TATAACATATGCTTGAAGTAGAAGATAGCAATTAATTTTACCCAGCTTATACAGAAC 606
Db 547 TATAACATATGCTTGAAGTAGAAGATAGCAATTAATTTTACCCAGCTTATACAGAAC 606
QY 481 ttatttcagttataaacaatgagccacaatagaagaatgcataatgagatgtagacct 540
607 TTATTTTCAGTTATTAACAATGAGCCACAATAGAAAGTCCATATGACATGTTAGACTT 666
Db 607 TTATTTTCAGTTATTAACAATGAGCCACAATAGAAAGTCCATATGACATGTTAGACTT 666
QY 541 atgagctctatatttgaagtgataacagtgctcagaagcttggatagagtttca 600
667 ATGAGCTCTATTTATTTGGAAGTGATACACTGCTCAGAGCTTTTGGATAGAGTTTAA 726
Db 667 ATGAGCTCTATTTATTTGGAAGTGATACACTGCTCAGAGCTTTTGGATAGAGTTTAA 726
QY 601 gtaaatctgtagactgctcacaagaatttaacaagaagcatatgattggcaaggtc 660
727 GTAATCTGCTAGCTGCTCATTAAGAATTTAAAGCAAGCATATGATTGGCAAGGCT 786
Db 727 GTAATCTGCTAGCTGCTCATTAAGAATTTAAAGCAAGCATATGATTGGCAAGGCT 786

| | | | | | | | |
|---|----|-------|------|--------|----|----------|---------------------|
| C | 9 | 443 | 10.6 | 72157 | 64 | AC016449 | AC016449 Homo sapi |
| C | 10 | 389.8 | 9.3 | 54398 | 79 | AC068224 | AC068224 Homo sapi |
| C | 11 | 232.2 | 5.6 | 2212 | 83 | AK021757 | AK021757 Homo sapi |
| C | 12 | 207.4 | 5.0 | 54398 | 73 | AC068224 | AC068224 Homo sapi |
| C | 13 | 206.8 | 5.0 | 113704 | 92 | HS267P19 | 275889 Human DNA s |
| C | 14 | 206.8 | 5.0 | 163642 | 79 | AL353724 | AL353724 Homo sapi |
| C | 15 | 179.6 | 4.3 | 193880 | 80 | AL358892 | AL358892 Mus muscu |
| C | 16 | 179.6 | 4.3 | 340969 | 81 | AL512630 | AL512630 Mus muscu |
| C | 17 | 163 | 3.9 | 163642 | 79 | AL353724 | AL353724 Homo sapi |
| C | 18 | 131 | 3.1 | 145450 | 73 | AC068352 | AC068352 Homo sapi |
| C | 19 | 131 | 3.1 | 160962 | 67 | AC022862 | AC022862 Homo sapi |
| C | 20 | 131 | 3.1 | 167671 | 77 | AC090081 | AC090081 Homo sapi |
| C | 21 | 124 | 3.0 | 69740 | 77 | AC087748 | AC087748 Homo sapi |
| C | 22 | 121.4 | 2.9 | 145450 | 73 | AC068352 | AC068352 Homo sapi |
| C | 23 | 121.4 | 2.9 | 162775 | 67 | AC023006 | AC023006 Homo sapi |
| C | 24 | 119.2 | 2.9 | 173613 | 4 | AC007475 | AC007475 Drosophill |
| C | 25 | 119.2 | 2.9 | 192763 | 4 | AC007474 | AC007474 Drosophill |
| C | 26 | 119.2 | 2.9 | 194634 | 65 | AC020286 | AC020286 Drosophill |
| C | 27 | 119.2 | 2.9 | 262731 | 5 | AE003823 | AE003823 Drosophill |
| C | 28 | 115.4 | 2.8 | 69740 | 77 | AC087748 | AC087748 Homo sapi |
| C | 29 | 108 | 2.6 | 1489 | 89 | AK023592 | AK023592 Homo sapi |
| C | 30 | 106 | 2.5 | 72157 | 64 | AC016449 | AC016449 Homo sapi |
| C | 31 | 99.8 | 2.4 | 7218 | 10 | 166494 | Sequence 14 |
| C | 32 | 98.6 | 2.4 | 162775 | 67 | AC023006 | AC023006 Homo sapi |
| C | 33 | 87.6 | 2.1 | 167671 | 77 | AC090081 | AC090081 Homo sapi |
| C | 34 | 83.8 | 2.0 | 160962 | 67 | AC022862 | AC022862 Homo sapi |
| C | 35 | 70.8 | 1.7 | 978 | 53 | CNS06PJK | AL409446 T7 end of |
| C | 36 | 68.6 | 1.6 | 156550 | 64 | AC015830 | AC015830 Homo sapi |
| C | 37 | 68 | 1.6 | 147146 | 83 | AP003316 | AP003316 Oryza sat |
| C | 38 | 67.6 | 1.6 | 66288 | 78 | AC090825 | AC090825 Homo sapi |
| C | 39 | 66.4 | 1.6 | 886 | 53 | CNS075DF | AL429961 Clome BA0 |
| C | 40 | 65.6 | 1.6 | 153665 | 78 | AL138818 | AL138818 Homo sapi |
| C | 41 | 65.4 | 1.6 | 23502 | 6 | CELZK355 | AC006832 Caenorhab |
| C | 42 | 65.2 | 1.6 | 74119 | 71 | AC036177 | AC036177 Homo sapi |
| C | 43 | 64.8 | 1.5 | 42839 | 70 | AC027282 | AC027282 Homo sapi |
| C | 44 | 64.6 | 1.5 | 68844 | 70 | AC027155 | AC027155 Mus muscu |
| C | 45 | 64.6 | 1.5 | 81120 | 67 | AC022851 | AC022851 Homo sapi |

ALIGNMENTS

RESULT 1
 HS26H231 7444 bp mRNA PRI 25-JAN-2000
 DEFINITION Novel human gene mapping to chromosome 13.
 ACCESSION AL137201
 VERSION AL137201.1 GI:6759511

SOURCE

ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS Rhodes,S. and Huckle,E.
 TITLE Direct Submission
 JOURNAL Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk
 COMMENT This cDNA sequence was assembled from public domain ESTs and single
 pass sequencing reads from expressed DNA templates, aligned to the
 genomic DNA sequence from the bacterial clones 26H23 (Z84467),
 267P19 (Z75889) and 49J10 (Z84572).
 The EST sequences listed match this sequence with an identity of at
 least 95% between the coordinates shown.
 Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene
 Sanger Centre name: 26H23 Cl3.1.

FEATURES

source location/Qualifiers
 1..7444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="13q12.3"
 1..107
 exon

| | | |
|--------------|---|------------|
| exon | /number=1 | 108..234 |
| exon | /number=2 | 127..4470 |
| CDS | /codon_start=1 | |
| | /product="hypothetical protein" | |
| | /protein_id="CAB69911.1" | |
| | /db_xref="GI:6759512" | |
| | /translation="MAHSTRNDGKITPPGVKIEISDKISKEEMVRRLKMKVTFMD
MDQSEEEKLEYLNALALASDFELKHPDKDVRLLVACCLADIFRIVAPDPYSPDK
LKDIFMFTROLKGLDPEKSPQPNFYFLLENIMVSKVNICFELSDSEITETQXYRT
LFSVYNNGNHNVHMYHNDLMSSTICEEDTYSQELDLYVNLVPAHNNLKQATDIA
KALKRTAOALEPTITNENOVLMIGKTSISDSEHVDLLELYNIDSHLLSVLPQ
LEFKLSNDNEERLOVVKLAKMFCAKSELSAKPKMOCYLGKFNIDHVPTRLEC
VFASHCLNNHDLADLLEYLKRSHDEEELARHDYIVSYIYAKKDDLLVNDHLNF
VREPTLDKRWVRKEMAMGLAOYKKVLAQSAAGDAKQAMIKDKLHLYYONSID
DLIYVERIEAOYVMPHNLFTREPMKCLYIYATDILNKKALNEMKCONLLRHQVD
LIDLITKPKTDASVAKAITSKVAVYTRNLIPDEKADDEPKKFTOVLEDEKTRKOLEV
VSPYSCSKOAEQCVREITRKLGPNKQPTNPPELMITKELLERAPVHIDTESISALIKO
VNSKIDGTADDEEGVPYDQAIKAGLILKVLSTFHPISFSAEFESILACLKMDDE
KVAEALQIFKNTGSKIEDPPIRSALLPVLLHNSKSGPQAKYAIHCIAITFSK
ETQFAOITPEPLHKSILDEPHILPLVTIGHIALADPQAPAPLKSIVAFYVLDL
MNDRLRGKTKTKIWPDEEVSPEPMVKIOAIKMYRMILGKNNHNSKGTSTRLITF
ILHSDGDLTBOGKISKPPMSRLRLMAGSAIYKLAQEPYHEITITEOQLDALINDE
CYQVQVFAQKLRKLSRLRPLEMAYICALAKDPVERERHARHOCIVKNNINAREY
LKQNAVSEKLSLLEPYVPTTHLLAHADPYVVDIEDLKVKDECLMFLVLELMA
KNENSHAFIRKMWENIKOTKDAQDPDAKMKELTYVCDAVAMNIMSKSTYSLESP
KDPVLPAREFTOPDKNFNTNKLILPEPKSPFTPEKPKPTTVYLGAVNKLPSAGKSO
TSSRMETVSNASSSSNPSPGRIKGRLDSEMOHSEMDYTMSSPLPGKRSKRSQSD
DIYRELEKPRGRKTPYTOBEKLGMDLTKYLVQEQPKGSQSRKRGHARSBEDQ
QMPERKRLKEEDILENEDQNSPPKGRKGRPKPLGGGTPEEPTMKSKGSKKSG
PPAPEEEEEEEOGSCGTEKSKOHRSRRAGQAPESPESALISTSTQKQGRP
SKTSPSQPKKRVYGRSKQOATKNDSEEDVDFQSSPVDDIPQEEETEVEEVSTVN
VRRRSKRRRR" | |
| exon | /number=3 | 235..438 |
| exon | /number=4 | 439..525 |
| exon | /number=5 | 526..623 |
| exon | /number=6 | 624..750 |
| exon | /number=7 | 751..831 |
| exon | /number=8 | 832..972 |
| exon | /number=9 | 973..1088 |
| exon | /number=10 | 1089..1183 |
| exon | /number=11 | 1184..1329 |
| exon | /number=12 | 1330..1481 |
| exon | /number=13 | 1482..1595 |
| exon | /number=14 | 1596..1677 |
| exon | /number=15 | 1678..1726 |
| exon | /number=16 | 1727..1866 |
| misc_feature | /note="matches EST AA687846 from clone IMAGE:1219676" | 1866..2046 |
| misc_feature | /note="matches EST AA687776 from clone IMAGE:1219628" | 1866..1941 |
| exon | /number=17 | 1867..1982 |
| exon | /number=18 | 1983..2088 |
| exon | /number=19 | 2089..2249 |
| exon | /number=20 | 2250..2373 |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 2001, 20:11:26 ; Search time 12186.1 Seconds
(without alignments)
5296.744 Million cell updates/sec

Title: US-09-512-581-3
Perfect score: 4173
Sequence: 1 atgctcatcaataaagactag.....cacacccaaaaaaatgtg 4173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pi1:*
- 13: gb_pi2:*
- 14: gb_pi3:*
- 15: gb_pi4:*
- 16: gb_ba1:*
- 17: gb_ba2:*
- 18: gb_fun:*
- 19: gb_hum:*
- 20: gb_hvg:*
- 21: gb_hvg_in1:*
- 22: gb_hvg_in2:*
- 23: gb_hvg_in3:*
- 24: gb_hvg_in4:*
- 25: gb_hvg_in5:*
- 26: gb_hvg_in6:*
- 27: gb_hvg_in7:*
- 28: gb_hvg_in8:*
- 29: gb_hvg_in9:*
- 30: gb_hvg_in10:*
- 31: gb_hvg_in11:*
- 32: gb_hvg_in12:*
- 33: gb_hvg_in13:*
- 34: gb_hvg_in14:*
- 35: gb_hvg_in15:*
- 36: gb_hvg_in16:*
- 37: gb_hvg_in17:*
- 38: gb_hvg_in18:*
- 39: gb_hvg_in19:*
- 40: gb_hvg_in20:*
- 41: gb_hvg_in21:*
- 42: gb_hvg_in22:*
- 43: gb_hvg_in23:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pi1:*

48: em_pi2:*

49: em_pi3:*

50: em_pi4:*

51: em_pi5:*

52: em_pi6:*

53: em_pi7:*

54: em_pi8:*

55: em_pi9:*

56: em_pi10:*

57: em_pi11:*

58: em_pi12:*

59: em_pi13:*

60: em_pi14:*

61: em_pi15:*

62: em_pi16:*

63: em_pi17:*

64: em_pi18:*

65: em_pi19:*

66: em_pi20:*

67: em_pi21:*

68: em_pi22:*

69: em_pi23:*

70: em_pi24:*

71: em_pi25:*

72: em_pi26:*

73: em_pi27:*

74: em_pi28:*

75: em_pi29:*

76: em_pi30:*

77: em_pi31:*

78: em_pi32:*

79: em_pi33:*

80: em_pi34:*

81: em_pi35:*

82: em_pi36:*

83: em_pi37:*

84: em_pi38:*

85: em_pi39:*

86: em_pi40:*

87: em_pi41:*

88: em_pi42:*

89: em_pi43:*

90: em_pi44:*

91: em_pi45:*

92: em_pi46:*

93: em_pi47:*

94: em_pi48:*

95: em_pi49:*

96: em_pi50:*

97: em_pi51:*

98: em_pi52:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 4165.6 | 99.8 | 7444 | 92 | HS26H231 |
| 2 | 4141 | 99.2 | 5253 | 97 | HS095825 |
| 3 | 4073 | 97.6 | 5110 | 85 | AB023196 |
| 4 | 1754.4 | 42.0 | 1852 | 97 | HS050533 |
| 5 | 1041.4 | 25.0 | 5177 | 85 | AB014548 |
| 6 | 447 | 10.7 | 2079 | 89 | AK026889 |
| 7 | 447 | 10.7 | 137246 | 92 | HS49J10 |
| 8 | 447 | 10.7 | 168487 | 89 | AL138820 |

Qy 560 RKQLEVLVS--PTCS-----CKQAGCVREITKKGNGPKQPTNPFLMIKFLLERIA 609
 Db 594 KSKNVAVSVEPTSSGVRSSRTLMKKDCGKRLNKQVEKTRGKN-----LRSLK 644
 Qy 610 PVHIDTESISALIKOVNKSIDGTADDEDE 638
 Db 645 ELNAETDRTA---EEQEVSLAESDDRSE 670

RESULT 14
 Q9S9P0 PRELIMINARY; PRT; 990 AA.
 ID Q9S9P0;
 AC Q9S9P0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE T24D18.4.
 GN T24D18.4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A..
 RC STRAIN=CV. COLUMBIA;
 RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,
 RA Toriumi M., Chin C., Chiu J., Choi E., Chung M., Gonzalez A.,
 RA Hwang B., Koo T., Li J., Liu A., Vaysberg M., Altafi H., Brooks S.,
 RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,
 RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RA "the sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A..
 RP STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RC Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN SEQUENCE FROM N.A..
 RP STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RC Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC010924; AAF18491.1;
 SQ SEQUENCE 990 AA; 109594 MW; 47854068D1B444FB CRC64;

Query Match 3.78; Score 264; DB 10; Length 990;
 Best Local Similarity 17.5%; Pred. No. 7.6e-08;
 Matches 241; Conservative 188; Mismatches 440; Indels 506; Gaps 50;

Qy 43 MDMDQDEERELNLAHLASOFFLKHCKGVRLVACCLADIFRIYAPEAPYSPDK 102
 Db 47 VEQDLSSVQKALPPHRYALVSAD-LRNPSDVRSVSVCLTEIMRTAPEAPY-NDEQ 104
 Qy 103 LKIDFMFTITROLGLEDTKSPQFNRYFYLLLENIAWKSYNICFELEDSNEIFTQLYRTL 162
 Db 105 MKDIFQVTEAFELADASSRYKAEVILETVAKVRSSLVMDLE-CDDLVLVEMFORFL 163
 Qy 163 SVINNHQKVVHMHVMDLMSIIICEGTVSOELLDVTLVNLVPAHKNLKNQAYDLAKALL 222
 Db 164 KIIRPDHPQLVLSMETIMITVIDESEVPMDLLEILLTTTVKKDSQDVSPAALTILVEKVL 223
 Qy 223 KRTAQAEIPIYITTFNVMVLGKTSISLSEHVFDLILELYNIDSHLLSVLPOLFE--- 279
 Db 224 SSCTCKLQPCIM-----EALKSSGTS-----LDMS-----PVSSICSEFATT 263
 Qy 280 -----KLKSDNDEERL---QV---KLLAKMFGAKDSELASQNKPLMOCYLGFRNDIHVP 328
 Db 264 QAHNDVVKPDNEADEKISEGVVPVDSLEDKLNLGLSRKGRSRKSRGGRANGD--- 320

Qy 329 IRLCEVCFASHCLMNHDPDLAKDLTEYLKVRSHDPPEEATRHDVIVSIVTAARKDILLVNDH 388
 Db 321 ---EKVITANEGLSESTD-AETASGSTRKRGWKPKSLMNPBEGYSFKTSSKKV----- 370
 Qy 389 LLNFVRETLDKRVRKEAMMGLAQIYKYVALOSAAACKDAKQIAWIKDKLLHIYQNS 448
 Db 371 -----QEKELG-----DSSLGKVAARKVP----- 389
 Qy 449 IDRLLLVERIFAQYVMVPHNLETTERMKCLYYLYATLDLNAVKALNMWKCQNLLRHQVKD 508
 Db 390 -----LPSKVGQTNQ-----SVVISLSSSGRARTGSRKRSR- 420
 Qy 509 LLDLIKQPKTDASVKAIFSKVMVITRNLPDQKGAQDFMKKFTQVLEDDK---IRKQLEVL 566
 Db 421 ----TKMEETHDVSSVATQ-----PAKKQT-VKKTNPAREDITKSNVKKHEDGI 465
 Qy 567 VSPTCSCQOA-EGCVREITKKGNGPKQPTNPFELE-MIKELLERIAPVHIDTESISALI 622
 Db 466 KTGSSKKEKADNGLAKTSARK-----PLAETMMVKPSGKKL--VHSDAK----- 508
 Qy 623 QVNVKSIDGTADDEGVPTDQAIRAGLELLKVLSTFTHPISFHSAETFESLIACLKMDDE 682
 Db 509 ---KKNSEGASMD---TPIQSSKS-----KKKDS 532
 Qy 583 KVAEAAALQIFKNTGSKTEEDPFHRSALLPVLHHSKKGPPRQAKYAIHCHIAIFSSRET 742
 Db 533 RATTPTAK-----KSEQAKSHPKMKR-----IAGEEV 560
 Qy 743 QFAQIFEPHLKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPKWSWATFIVKDLLMNDR 802
 Db 561 E-----SNTNEL-----GEE 570
 Qy 803 LPGKTTKLWYDPEVSPETVMVKIQAIKMVMRWLLGMKNHNSKSGTSTLRLLTTLHSDG 862
 Db 571 LVGRVNVNWPDLKKFEGVYKSVCRVKKM-----HQ-----VTYSOG 608
 Qy 863 DLTEQKISKPDMSRLRLAAGSAIVKLAQEPYHEITLQYQLCALAINDECTQVRQVF 922
 Db 609 DV-BELNLKK---ERFKIIEKSSASEDKEDDLLESTPLSAF----- 646
 Qy 923 AQLHKLGLSLRLPLEYMAICALCAKADPVKERRAHQCLVKNINVRREYLKQHAHVSEK 982
 Db 647 -----SSPEVRSSMQTMKKD-----IQREKSKKRKIVSKN 661
 Qy 983 LLSLLPEYVVPYTIHLLAHPDYVKVQDIEQLKDVKECLFVLEILMAKNENNENSHAFIRK 1042
 Db 662 -----VEP-----SSPEVRSSMQTMKKD-----S 682
 Qy 1043 MVENIKQTKDAQGPDADAKMNEKLYTVCVAMNIMSKSTTYSLESKPDVLPAREFTOPD 1102
 Db 683 VTDISIKQTKRTKGALKAVSNPE-----ESTGKNLKSLLK-----KLNGEPD 723
 Qy 1103 KNFSNTKNYLPENKSFPTGPKPTTNVLGAVNKPPLSAGKOSOTKSSRMETVSNASSS 1162
 Db 724 KTRGT-----CKQ-----KVTQAMHRKIEK----- 745
 Qy 1163 NPSPGRIKGRDLSMDHSENEEDYTMSSPLPGKSKDRDSDLVRSLELPRGRKKTTPV 1222
 Db 746 -----DCDEQEDLETKEEDSLK-LGKESDA--EPDRMEDHQELPENHNVEK 790
 Qy 1223 TEQELGCMDDTLKLVQEQPKGQSRSPKRGHTASEDEQOWPEKRLKEDILE--NEDE 1280
 Db 791 TDGEE-----QEAKEPTAESKTNGEENAEPTDCKEHSKLUKEPNAEPKSDGE 839
 Qy 1281 QNSPPKK-----GKRGRPPKPLGGTTPKEEPTMTKS-----KKGSKKSGPPAPEEE 1328
 Db 840 EQEAAKPEANELKTDGQENQEAKEKELTAERKTDDEEHKVADEVEOKSQKETNVEPAESEE 899
 Qy 1329 EEEQSGNETQKSKSHRVSRRRAQQAESPESSAESTQSTQGRGRPKSKTPS 1383
 Db 900 QKSVEEPNAEPKTK-----VEEKESAKEQTADTKLIEKEDMSKTKGEBIDKETYS 949


```
Db 61 GRPKPLGGTPEPTMKTSKSGKSGPPAPPEEEEEERQSGNTEQKSKSKOHRVSR 120
Qy 1351 RAQR 1355
Db 121 RAQR 125

RESULT 11
ID 094076 PRELIMINARY; PRT; 1506 AA.
AC 094076;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE BMD PROTEIN.
GN BMD.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93387663; PubMed=8375649;
RX Denison S.H., Kaefer E., May G.S.;
RT "Mutation in the bmd gene of Aspergillus nidulans confers a
RT conditional mitotic block and sensitivity to DNA damaging agents.";
RL Genetics 134:1085-1096(1993).
CC -!- FUNCTION: REQUIRED FOR PROGRESSION THROUGH MITOSIS AND ALSO PLAYS
CC A ROLE IN REPAIR OF DNA DAMAGE.
CC -!- SUBCELLULAR LOCATION: KINETOCORE OR MITOTIC SPINDLE (POTENTIAL).
CC -!- SIMILARITY: SOME, TO B-ZIP PROTEINS.
DR EMBL; L03200; AAA03063.1; -
KW Cell cycle; Mitosis; DNA-binding; DNA repair.
SQ SEQUENCE 1506 AA; 166270 MW; F50B0893EE767579 CRC64;

Query Match 8.8%; Score 632.5; DB 3; Length 1506;
Best Local Similarity 19.9%; Pred. No. 4.5e-30;
Matches 308; Conservative 298; Mismatches 626; Indels 319; Gaps 53;

Qy 30 EMVRLKVVTFMDQDSEEEKLYLNLAHLASDFLKHGPKDVRLLVACCLADIFR 89
Db 47 DLLQRLQTLAQLPNVEQE-EIERESLRKVSOELATAQLLAHKDGVRAWATCCIVDLR 105
Qy 90 IYAPEAPYTPDKLDIFMTIRO-LKLEDTKSPQFNRYFYLLNIAMVSKYNTICFELE 148
Db 106 LCAPDAPETA-NQLKDIFTCTIVSSIIIPALGDPSPYNAQHIYVLNLAEVKASIVLMTDLD 164
Qy 149 DSNEIFTQLYRTLSVINNGHN-----QKVHMHVYDLMASSICEGDTVSQELLDIVL 200
Db 165 HPOTLIVPLFISCFDIVAGSAKASTGEPVAKNVEYDMTRLLVTVIDESPVLAPDVVIV 224
Qy 201 VNLV-----PAKKNLQK-----AYDLAKALLKRTAQALEPYIT 234
Db 225 AQFLRVDPRLVDGCKKCKEPTQVDEKQETLLLLKDYPAAYNMAKICACPERMTSHS 284
Qy 235 TFFNOVL-----LKTSTISDSEHFVD-----LILELYNIDSHLLL 271
Db 285 QYFNNVIIDASATGTQGPSQARRTNLDDSEEGEDIKLSKAHRLIRELWRACPDVLQ 344
Qy 272 SVLPQLEFKASNDNEERLQVVKLAKMFGAKDSLASQNKPL----- 314
Db 345 NVIPOLEAELSAESVSLRLATQTIGDL--TSGTGAGVAGPPPLPMDPAPVQVQLDDYAR 402
Qy 315 -----WQCYLGRFNDITHVIRLECVKFAASHCLMN----- 343
Db 403 SIPOPNVLLMPFAPKPPFQAHSAYDSFLSRLDKSASRWRTAIGRIILTSAGGSL 462
Qy 344 HPDLAKDLTEYLVKVSHPDEBAIRHDVITSVITAAKDI-----LLVNDHLLNVR 394
Db 463 SDNEQTLTILHSLMLRDADERVRLAAEAVGTFGLSHVKNKLVGSGVSTQDSLLFLA 522
Qy 395 ERTLDKRRVRKEAMWGLAQIYKYKVALQSAGKAAQIAWI-----KDKLLHIYQNSID 450
```

```
Db 523 ERVDRKSQVREHATKYL-----RAMAVASGDIERSHQVTPLLKEAPSRILDAYTNDPE 579
Qy 451 DRLIVERIFAQYMP-----HNLETT-----RMKCLY 479
Db 580 IHVSIDRAMPEILLPLSYPIKPKLSRSSSSQSLKDSQAAPESADVDRIVRILT 639
Qy 480 LYATLDLNAVKALEMMKQNLRRHQVLDLILQ-----PKTDASVKAIFSVM-VI 532
Db 640 LVGGLDEKAKKVFAMQKROVSLRTAVTVYLQACEEYNGVMEKNKQDKAQLTKIVDAL 699
Qy 533 TRNLPDPCGAODFMKKFTQVLEDEKIRKQLEVLVSPFCSCQKQABGCYREITKTLGNPKQ 592
Db 700 AKTPDPARTSADLWKFAK--HRRGYOLIRFAMAAVSDYRTVIKAIKELARRL---OS 754
Qy 593 PTNPFL-EMIFLERIAPVHIDTESALIKQVKNKSDGTADDEGVPTDQAIRAGLE 651
Db 755 SNNILHETLTLTYRCSSIVFNRSHPA-IMSISRS-----DENGLAAPAH-----E 801
Qy 652 LLKVLSTPHPISTFSAETTESLL--ACLKMD-----DEKVAEALQIFRWNG 696
Db 802 MLKEISSLNP-----EVLEAQVQEICKLEAQAPKATTVSAAAGTEELKACSGPAKLP 855
Qy 697 SKIEDPPIRSALLPVLHHSKKGPPROAKYAIHICHAIFSSRETQFAQIFELHKSLD 756
Db 856 SKLPKERKFFQALVDYALHSPS-----PRAAKHAVLILMAV-TDKKNMYAKDLVEKVCSC 910
Qy 757 PSNLEHLITPLVTGHIALAPDOFAAPKSWATFIVKLDLMDRLPGKTKTKLWVDE 816
Db 911 TYDSERFLTKATLSQNLAPRE-ADESDAIKISVNOILLTNRSPTPNSGYFW--SD 967
Qy 817 EVSPETVVKIOAIKMMVRLMGMNNHKSSTST-----LRLTTILHSDGDLT 865
Db 968 QVDDTAKEWALKIIVNRL-----RAKDGSDSDDFRAHAEPVYDTLNLKLVNSGELS 1021
Qy 866 EGGKISKPDMSRLRAGSAIVKL-AQEPYHEIITLQYQOLCALAINDECYQVROVFAQ 924
Db 1022 KKTPTATQKSLRLLAANSLLKLCSSHALCEQLTPODFNSIALVAODPLPEVRSGFIN 1081
Qy 925 KLHGLSL-LRLPLEYMAICALCARDPVKERRAHARQCLVKNINVRREYLKQHA----- 977
Db 1082 QLKKKLVQDTRLGARWVIVPYLLAFEP---QVGLKQSTLTWLSRAAFPSQOTNGKKE 1137
Qy 978 --AVSEKLLSLPEVVPYTTIHLAHDPPY--VKVQDTEQLKDVKECLWFVLEILMARNE 1033
Db 1138 KQTYMEALFSRL-----LSLLAYHPDYPADLDESTKLDLTDFAFYILFYLSAVAN 1189
Qy 1034 NNSHAFIRKVVENIKQTKDAQOPDDAKMNEKLYTVCDVAMNINLSKTYTSLESKPQVPL 1093
Db 1190 EHNLSLIFHIAQVRKQARQDGTIKSD-EMSRRLHTLSDLAQATIRFADVYQQ----- 1241
Qy 1094 PARFTQPDKNFSNTKNYLPPEMKSFPTPGKPTTNVLGAVNKPLSSAGKOSQTKSS--- 1150
Db 1242 --RRFGGAGGVNLLQTY-----PGK--VGVPSSIFAPMGSHREAEVADTTFL 1286
Qy 1151 -----RMETVSNAS-SSSNPSPGRIRKRLDSSSMHSENEEDYTMSSPLCKKSKDR-D 1202
Db 1287 PEDAEIDLRLIVRATMTKNGSQAGAKRKRTDSTQEPSRDANAAKKARKNSESRRKS 1346
Qy 1203 DSDLVRSLEKPRGKKT--PVTEQEELGLMDDLTKLVQEQPKGQSRSRGRHGTASESD 1260
Db 1347 GSSAVGFKMPKRSKTKTKGDDWSSDGEAGNVASSSATRRRSNRGS-ASRISYADPOSD 1405
Qy 1261 E--QWPEEKRLKEDIENEQNSPPKKGRGRPPKPLGGGTPEKPTMTKTSKSGSKK 1318
Db 1406 EDDMEMDELNQAQDEDEGDEQ-----AKDIENGSDLSE---LSEADSNWL 1448
Qy 1319 SGP-----PAPPEEEEEERQSGNTEQKSKS-----KQHRVSRRAQRAESP 1359
Db 1449 EPEEDDGPSEKEDQPDQKQDDDDAQAPSPVPVASKAKVPCAMKATLP 1499
```

RESULT 12

| | | | |
|---|--|--|--------------|
| Db | 1074 | KLOKYLADGKLSRYYTIFLAFPSVEFK-----NRVETWIRSRARHFONLK | 1122 |
| Qy | 985 | SLLPXYVVPYTHLLAHPDYYVKVODIEOLKDVKECLMFVLEILMAKNENNNSHAFIRKMV | 1044 |
| Db | 1123 | QPVLAEIMARLISLAAHPDY--SNEVEYLVDPHAR--YILNYIVLVGTESNLGLIYKYA | 1177 |
| Qy | 1045 | ENTKQTKDAQGPDPAKMKNEKLVTCOVAMNII---MSKSTTYYLESKPDVPLPARFT-- | 1099 |
| Db | 1178 | ERVKQTDQGLNPN---SDAHRVLSDLAQSIIRKWKQEKVWAFAPCGKVGPLGLLYTAL | 1233 |
| Qy | 1100 | -OPDKNFSNWKYLLPP-----EMKSFTTGPCPKTTNVLGAVNKPLSSAG | 1142 |
| Db | 1234 | QSHSEAQAIAEKSYLPEGLDEKLDELLRAMDRKKRKRTSAATGTTNRDGHKSKVKKSSG | 1293 |
| Qy | 1143 | KQS---QTKSSRMETVSNASSNPSGRIKGRLDSEMDHSEMEDYTMSSPLPGKKSD | 1199 |
| Db | 1294 | DROGDREAMSSR--KTATAKSKTKTAK-----KRSDDDWSPCAAYP--ESE | 1336 |
| Qy | 1200 | KRDSDLVRSLEKPRGRKKTPTVTEQ-----EKLGMDDLTKLVOEQPK | 1244 |
| Db | 1337 | RQSS-----RRAANKAKGPVTERSDDDDEEMLEGVAEWYLDDEDDKEGDDEGS | 1389 |
| Qy | 1245 | GSQ-----RSRKGHTASSD--EQWPPEKRLKEDILENEDE-----QN | 1282 |
| Db | 1390 | GEAAPPRRTAAKASATPRTSSRREDKESSEEVNEEPEKEQEEEAADPEQEEAPTAAAKRT | 1449 |
| Qy | 1283 | SPPKKKGR-----GRPPKPLGGTGPKEPTMKTSKKSKKSGGPPAP | 1324 |
| Db | 1450 | SATAKGGKAASLPVNAASAASTRSGRGKAAAAAASAKDKDSDLSLSDVEMVDEPEA | 1509 |
| Qy | 1325 | E-----EEEEERQSGNTEQKSKSKQHRYSRRAQRAESPSSAISTEQST | 1370 |
| Db | 1510 | EAEAEAEAEVAEVEDEQEQAAPPARANGRKAKAAAPAKAKAAAPGKAKAADASA | 1569 |
| Qy | 1371 | PQKGRGRPSK 1380 | |
| Db | 1570 | PARRSGRSTR 1579 | |
| RESULT 10 | | | |
| ID | Q9H5N8 | PRELIMINARY; | PRT; 125 AA. |
| AC | Q9H5N8; | | |
| DT | 01-MAR-2001 (TrEMBLrel. 16, Created) | | |
| DT | 01-MAR-2001 (TrEMBLrel. 16, Last sequence update) | | |
| DT | 01-MAR-2001 (TrEMBLrel. 16, Last annotation update) | | |
| DE | CNDA: FLJ23236 FIS, CLONE COL00725. | | |
| OS | Homo sapiens (human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| NCBI_TaxId | 9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=COLON; | | |
| RA | Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., | | |
| RA | Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T., | | |
| RA | Tanaka T., Nakamura Y., Isogai T., Sugano S.; | | |
| RT | "NEDO human cDNA sequencing project."; | | |
| RL | Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AK026889; BAB15584.1; | | |
| SQ | SEQUENCE 125 AA; 14281 MW; 492809836D336565D3 CRC64; | | |
| Query Match 9.1%; Score 656; DB 4; Length 125; | | | |
| Best Local Similarity 100.0%; Pred.No. 5.4e-33; | | | |
| Matches 125; Conservative 0; Mismatches 4; Indels 0; Gaps | | | |
| Qy | 1231 | MDDLTKLVQEQPKGQSQRKRGHTASDEQOWPEERKLKEDIENEDEQNSPPKKGR | 1290 |
| Db | 1 | MDDLTKLVQEQPKGQSQRKRGHTASDEQOWPEERKLKEDIENEDEQNSPPKKGR | 60 |
| Qy | 1291 | GRPPKPLGGTGPKEPTMKTSKKSKKSGGPPAPEEEEEERQSGNTBQKSKQHRVSR | 1350 |

| | | | | |
|-----------------------|---------|--------------------|-------|-------------|
| Query Match | 9.1%; | Score 656; | DB 4; | Length 125; |
| Best Local Similarity | 100.0%; | Pred. No. 5.4e-33; | | |

1231 MDDLTKLVQEQPKGQSRSKRKRGHTASDEQEQWPEEKRLKEIDLENEDEQNSPPKKGKR 1390
y
b 1 MDDLTKLVQEQPKGQSRSKRKRGHTASDEQEQWPEEKRLKEIDLENEDEQNSPPKKGKR 60
y 1291 GRPKPLGGTPKPEEPTMTKSKGSKKKSGPPAPEEEEEEROSGNTQSKSKQKHVSR 1350

Query Match 11.2%; Score 807; DB 5; Length 1579;
Best Local Similarity 21.5%; Pred. No. 1.1e-40;
Matches 352; Conservative 282; Mismatches 674; Indels 326; Gaps 57;

Qy 11 GKITYPGVKESDKISKEEMVRRLKVVKTFTMDMDQDSEEE---KELYLNALHLASDF 67
Db 3 GDVYPOCCLPILNSNNVTOQIERLURKLFNCLHDKTNNSEVSPNRFARLQHLSDQC 62

Qy 68 FLKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLIDFMFITROLKLEDT--KSPQF 125
Db 63 FLDNSNTDFRILLSLCLANILRIFQPELPTSPVMDKEVYVLFRTMRGLGVDVTDSPKF 122

Qy 126 NRYFYLLENIAWVKYNICFELEDSENE-----IFTQLYRTLFVSV-INNGHQ----- 171
Db 123 KNYSLSVETME--KIIPPIIEMKDDHDEATPVFRALIKLILAPCGKGNQNLKEARL 180

Qy 172 -----KVHMHVMDLMSSTIICGDTVSQBELLTVLVNLVPAHKN 209
Db 181 LKIQDNDSDNDDEEDENAEKIRKSLIQATVITNLDVQNECLDVLVYHIINPORS 240

Qy 210 LNKQAYDLAKALK-----RTAAQIEPYITTFNFQVLMGLK-----TSISDLSEH 254
Db 241 NFAEARALAEIDIIRSCDNESTLANSIRSTMTAAAKE-----GKLPEEFELTGSNSRSK- 295

Qy 255 VFDLILELYNDSHLLSVLPQLEPKLSNDNEERLQVYVLLAKMFGAKDSELA--SQNKP 313
Db 296 FFEVLYLHYVSFVLVGAIGELFQWQSENEQYRKEATVYVGMLTRDKHCFQFGMDSDNP 355

Qy 314 LWQCVLGRFNDIHPVIRLECVKFASHCLM--NHPDLAKOLTEVLRVSHDPEAIRHDVIV 372
Db 356 TWSAFNLASIDQDVSREHFVQOQKDIILSNHSLRGQIINSLLSLVDLDDIRRDVVT 415

Qy 373 SIVTAAKDKILLNDHLLNFVREFTLRKWRVRKEDAMGLAQIYKYLQ----- 422
Db 416 GVTEVAKTKLEVISDKMLCAEAKMRKDKPKVRIQAIKRLMDLYNHVMTSSPPQPFSDKG 475

Qy 423 SAAGKDA-----AKQIAWIKDLHLHY-----YNSIDRLRLVERIFAQYM 463
Db 476 SCAPKDSKEATLSYTESKESVRFISTAVFNRYLRTORLPAYH-----DARATIERFYQLY 532

Qy 464 VPHNLLETERRKCLYLYATDLNVAKALNEMKQCONLLRHQVKDLDLIKQ--PKTDAS 521
Db 533 VPKYKPELRLVRLMADLFRNLDDICMLFGDIINSSQLRRAMIGILSOVQOFTMTTQS 592

Qy 522 VKAIFSKVWVITRNLPDPGKAQDFMKPE--TVLEDEK---IRKOLEVLVSPSCCKQAE 577
Db 593 SAQLKERIRRICQIPDQAILEKNMVFVQMAENDETFLNVKKLMSSEYTSSEENARTAG 652

Qy 578 GCVREITKLGKPNQPTNPFLEMIFKELLERAPVHIDFESISALIKQVNSIDGTADDED 637
Db 653 ALQATMEKKSISKAQOT-----VFRHFIDRIVPLSFVDVPTAKEIHLVSDTVCAKVDLKK 707

Qy 638 EGVPTDQAIRAGLELLKVLSETHPSFSAETFESLLACLKMDDEKVA--EAAL-----QIF 692
Db 708 W-----AENCFERDLSLKIFTDNFGVLFDAEQIIEIRSKILASEEPIAEALHVLSKIF 764

Qy 693 KNTGSKIEEDPHIRSAALLPVHLHKS-----KGPP-----RQAKYATHCTHA 735
Db 765 ANS-----HFRNKLENEATHEKKEFWLLGLKDLKDLVMREPELRSSCKLATRLLSF 815

Qy 736 IFSSKETQAFIPEFLHSLDPSNLEHLITPLVTIGHTALLAPDOFAAPKWSWATFI-- 793
Db 816 ILGKEKV--IEFFD-----DQIEQLISRLYIESOGAANAQVGLGIFRCDSIYILPQ 865

Qy 794 VKDLIMNDRL--PGKRTTKLWVPDEVS-----PE--TWVKIOAIKMMVRLWLLGM 839
Db 866 VMDVVESEKIGPMILTSWHGNDOPVEFNEMLHTEKQWPKYAMAKVYAANFAFKVLIVY 925

Qy 840 -----KNNHKSQGTSTLRLLTLTILHSDGLTEBQGIKSPDMSRLRLAAGSATVLAQ 891
Db 926 PLIPSTEDKRMEKAAQNFIDLLSEIEKKGDG--GGQCDCEQARLRATASGCLLKLAS 983

Qy 892 EPCYHEITLEQYQALCAINDEBQYQVQVFAQKLHKLHGLRLPLPLEYMAI-----CALC 946

Db 984 VITYRTKLINTHFKNMSYIITDEAYCVRLYYALHVKKGLSKNRLPIEFAACYGLVNLGLS 1043
Qy 947 AKDPKERRAHAROCVKVKNINVRREYLKQHAASVEKLLSL-----LPEVVPVYTHL 998
Db 1044 EEDGENKMDGFEKTCM-----NQAQAFGERNDNQTALLKLEGQRAIFCSSEVIAVYVL 1099
Qy 999 LAH-----DPDYVKVQDIEIOLKDVKECLMFVLEILMAKNENNSHAFIR 1041
Db 1100 LANYDKLEKVEGNANRNDSSBELEIKVANVNLSELQSLWLVIDSLKIAKCMQKVW-- 1157
Qy 1042 KMWENIKOTKDAQPDPAK-----NEKLYTVCDVAMNIM----- 1077
Db 1158 KVLKELKTCGDKSMRSDSLSTRLREHNEVNPDSGRNIKWLVCDLGITMMLYRAKLOM 1217
Qy 1078 -----SKSTYSLE-----SPKDPVLPAREFTOPDKNFNTKN-----YLPPEMK 1117
Db 1218 EDQEAKEAGFNLFQFYVCSPKDKADPSNVA--PDVLIINDEXNRNGRIPKGRVPHVSDLT 1276
Qy 1118 SFFTPGKPTTNVLGAVNKPILSSAGKQKQOTKSSRMETVSNASSSNPSSPGRIKRLDSS 1177
Db 1277 SEFTP--PPQGN-----ESTGSNSKSNASRRANVTGTRKRGAGTKTKRKSGGS 1325
Qy 1178 EMDHSENEEDYTWSSPLPGKSKDKRDDSDLVRSLEK-----PRGRKKTPYTEQEEKLM 1231
Db 1326 KISDDEGSDMDVKSPAVKTKRSKGEYDLPEDEDEMEVILPKRRGAAPDSTIASSSN 1385
Qy 1232 DDLTKLVQEQKP-----KGSQRKRGHTASEDEQOQWPEEKRLKLEDENEEDQN 1282
Db 1386 GSILKNGSGSPKKNRSRGVKGONSTKEAIDVDSDEM-----EETEKRDVNS 1437
Qy 1283 -----SP---PKGKRGRRPKPLGGGT--PKEPTMTKSKGKKKSGPPAPEEEEE 1331
Db 1438 LDNLIIISILDSESSGTRRSARTIATATITSSTPLVTPKTKITRKRSSEAVEVEDEE 1497
Qy 1332 RQSGNTEOKSKSKOHRVSR-----AQORAESPESSAI--ESTOSTPKGR--GRP- 1378
Db 1498 PEI--NDKRSPPKRVSGRRSAPTPTKNSKAPVSPKKTLIKETNGVSPKKNKYGILPM 1555
Qy 1379 -----SKTPSPS 1385
Db 1556 EEDDDSIGKTPKPT 1569

RESULT 9
QYUYV6
ID QYUYV6 PRELIMINARY; PRT; 1596 AA.
AC QYUYV6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SPO76 PROTEIN.
GN SPO76.
OS Sordaria macrospora.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Sordaria.
OX NCBI_TaxID=5147;
RN [1]
RP SEQUENCE FROM N.A.
RA van Heemst D., Poggeler S., Zickler D.;
RT "The SPO76/PS57/BLM gene involved in mitotic sister chromatid
RT cohesion is also needed for synaptonemal complex formation and
RT recombination in Sordaria macrospora."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ009934; CAB51808.1; -
SQ SEQUENCE 1596 AA; 176745 MW; B03FBB8C3563AEC2 CRC64;

Query Match 10.4%; Score 747.5; DB 3; Length 1596;
Best Local Similarity 20.7%; Pred. No. 4.7e-37;
Matches 338; Conservative 295; Mismatches 652; Indels 345; Gaps 51;
Qy 13 IYPPGVKVEISDKISKEEMVRRLKVVKTFTMDMDQDSEEEKELYLNALHLASDFLAKHP 72

| | | | | |
|--------|--|--|-------------|----------|
| Qy | 1046 | NIQTQDAQGDPAKAKNEKLYITVCDVAMNIMSKSTTYSLESPPKDPVLPARFFTQ--- | PDK | 11010 |
| Db | 1033 | SIKQSEV---TDATKSKNSHAICELGLSII----- | NHLTKQKPEDL | 1070 |
| Qy | 1104 | NFSNTKNYLPPEMKSFETPGKPKTTNVLGAVNKPQLSSAGKOSQTKSSRM--- | ETV--- | 1155 |
| Db | 1071 | QGEITPVSP---TLYKPE-----KVGGDKSQVGEELKWLUADETVLLHF | 1113 | |
| Qy | 1156 | -----SNASSSSNPSPGR---IKGRLDSSMDHSENYDTPSPPLPGKKSKDRDDSDL | 1206 | |
| Db | 1114 | RAULKLESHADASVIPQTSSENVWIDGESDQNEI----- | PL-GKIVER- | 1154 |
| Qy | 1207 | VRSELEKPRGRKKTPTVTEQEEKLGMDDLTATKLQVE-----QPKGSQRKRKGHTPASE | 1258 | |
| Db | 1155 | LRAQGTGTRKGKKNKSPVAEDENKNGNDVLAKKVREINLDHQLMDKPFESSNGHKHSPSE | 1214 | |
| Qy | 1259 | SDSQWPEEKRLKEDILENEDEQNSPP-KKGKRGRRPKPLGGGTGPKCEPTMKTSSKSGSK | 1317 | |
| Db | 1215 | RAETCQRDQGNKKNVGDATSVVVPKRRSSGHSPIKFSNSGPKVQ----- | LK | 1264 |
| Qy | 1318 | KSQPPAPPEEEEEEROSGNTQEKSKSHRVRRRAQRAESPSSAISTOS----- | 1369 | |
| Db | 1265 | ASEDELHLESDMDKNVLSHDNSDQEKMLSEISPRKRKLSLKKLITSDWALTQDVE | 1324 | |
| Qy | 1370 | TPQKGRGRPSKTPSPS---QPKNV | 1391 | |
| Db | 1325 | RSRSAGGDSKLSASGSMKRRKNV | 1349 | |
| RESULT | 8 | | | |
| Q9XTF6 | | | | |
| ID | Q9XTF6 | PRELIMINARY; | PRT; | 1579 AA. |
| AC | Q9XTF6; | | | |
| DT | 01-NOV-1999 (TrEMBLrel. 12, Created) | | | |
| DT | 01-NOV-1999 (TrEMBLrel. 12, Last sequence update) | | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last annotation update) | | | |
| DE | H38K22.1 PROTEIN. | | | |
| GN | H38K22.1. | | | |
| OS | Caenorhabditis elegans. | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; | | | |
| OC | Rhabditidae; Peloderinae; Caenorhabditis. | | | |
| OX | NCBI_TaxID=6239; | | | |
| [1] | | | | |
| RN | SEQUENCE FROM N.A. | | | |
| RA | Almscough R.; | | | |
| RL | Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases. | | | |
| [2] | | | | |
| RN | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=94150718; PubMed=7906398; | | | |
| RA | Wilson R., Almscough R., Anderson K., Baynes C., Berks M., | | | |
| RA | Barfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., | | | |
| RA | Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., | | | |
| RA | Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., | | | |
| RA | Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., | | | |
| RA | Lightning J., Lloyd C., McMurray J., Mortimore B., O'Callaghan M., | | | |
| RA | Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R., | | | |
| RA | Smalton S., Smith A., Sonnhammer E., Staden R., Sulston J., | | | |
| RA | Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., | | | |
| RA | Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.; | | | |
| RT | "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. | | | |
| RT | elegans"; | | | |
| RL | Nature 368:32-38(1994). | | | |
| [3] | | | | |
| RN | SEQUENCE FROM N.A. | | | |
| RA | Barlow K.; | | | |
| RL | Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases. | | | |
| DR | EMBL; Z50874; CAA90774.1; | | | |
| DR | EMBL; AL024499; CAA90774.1; JOINED. | | | |
| DR | EMBL; AL024499; CAA19710.1; -. | | | |
| DR | EMBL; Z50874; CAA19710.1; JOINED. | | | |
| SQ | SEQUENCE 1579 AA; 178885 MW; ADE29B7020BF1AC4 CRC64; | | | |

AC Q9V646;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE CG17509 PROTEIN.
 GN CG17509.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
 RA Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puril V., Reese M.G.,
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003823; AAF58593.1; -;
 DR FlyBase; FBgn0033676; CG17509.
 SQ SEQUENCE 1218 AA; 138887 MW; F4A6B7238A0801C2 CRC64;

Query Match 27.6%; Score 1982.5; DB 5; Length 1218;
 Best Local Similarity 37.3%; Pred. No. 1.5e-112;
 Matches 444; Conservative 247; Mismatches 421; Indels 77; Gaps 22;

QY 193 QELDTVLVNLVPAHKNLNKOAYDLAKALLKRTAQATEPVITTFNOVLGK--TSISD 250
 DB 183 VELDULILINIVEPYKNNKFAQLTEQLTKTGDALESTIKMFFNRLVMDPNTKLS- 241
 QY 251 LSEHVFDLILELYNIDSHLLSVLPQLEFKIKSNDNEERLQVVKVLLAKMFGAKDSELASQ 310
 DB 242 ITNKIYDIIVELARINAGLLCSVLPQLENKLLSTDAAERLKAITLLSRMFSEKDSOLAKK 301
 QY 311 NKPLWCYGLGRFNDIHVPRLCEVKFASHCLMHPDLAKDLTEYLKVRSHDPDEARHVD 370
 DB 302 YPNLLKIFFGFCDITEPVRIKCVQSSMHFLLNHPSLQHDITEKRLNRHDLDEVVRHEV 361
 QY 371 IVSIVTAAKKIDILLVND--HLLNFVRETLDKRWVRKKEAMGLAQIYKKYALQSAAGKD 428
 DB 362 VMAIVETAKRDFTLVLEAPDLLEIVRETLDDKKYIKRRDAMGLAYIYKRAICEPNDLST 421
 QY 429 AAK-QIAWIKDKLLHIYQNSIDRLIVERIFAQYVWPHNLETTMRMKCLYIYATLDLN 487
 DB 422 GLKVRVDWIKNKLHGYYKVGLEDRLIVERLLITCLVPYKLAPEERMKKLYHLGLDLDAN 481
 QY 488 AVKALNMKQCNLLRHQVKKDLDLKQPK-TDASVKAIFSKVMVITRNLPDPGKAQDFM 546
 DB 482 ATKAFVLOKQNMKTRNTVSDWIKLHHSKEFTPRVLSQLSAKQANAKLLPDLKAAEYL 541
 QY 547 KKTQVLEDEKIRKOLEVLVSPSCCKQAEKGVRETTKKLGNPKOPTNPFLEMIKFLLE 606
 DB 542 TQFSNNLRKDAQLLRCINIVLRDVSRECADTNGVLLKKLGAHVQ-SNIYYNTVMKMLE 600
 QY 607 RIAPVHIDTESIALIKQVKNKSIDGTADDEGVPTDQAIRAGLELKLVLSTFHPISFHS 666
 DB 601 RVASVMVDKESIGVLISLIEQCEKSGMCEIGISAQAGEGRGLKLLAMLISYVFSAHFT 660
 QY 667 AETFESLLACLKDDKVAEAAQIFKNTG---SKIEEDPHIRSAALLPVLLHHSKKGPP 723
 DB 661 DTSRLHLISLSLEYQDYVAPVLKTLTHLGRYQPLIDDPFPAILEDAPYKDFALIGTP 720
 QY 724 RQAKYAIHCH-----AIFSSKETO-----FAQIFELHSLDPSNLEHL 763
 DB 721 KQAKHVRCLIFVNSQSSASTDGTATSGAGSASTTQTVHPFNEIETRLKLRP-NCEHQ 779
 QY 764 ITPLVITGHIALAPDOFAAPKSWATFIVKDLLMNDRLPGKTKTL-----WVPDEEV 818
 DB 780 RTKIVTLGHITAYNMQPQAFITPIKNMIAIRIVKELLIOE-VPAQRDYELPEDSDWCAQEL 838
 QY 819 SPETMKIQAIAKMMVRLKMKNNHKSSTGLTLLTTLHSDGLTEQGIKSPDMSRL 878
 DB 839 PPDTLCKLDALKAMARWLLGLRTD-EHAAQKTFRMLAAAFVNRQDGLLQNRLCGAESWL 897
 QY 879 RLAAAGSAIVKLAQPCVHEITILEQYQCALAINDECYQVQVFAOKLHGLSLRL--- 934
 DB 898 RLGAACALAKVCQKGVGDYQSAEQYLQSLQMDADPVPEVREIFARKLHGLSLRSLPNC 957
 QY 935 LPLEYMAICALKADPKYKERRAHARQCLVKNNINVRREYLKQHAAS-----EKLLSLP 988
 DB 958 LPDLFMGLYVLAGEITERKQLDLVRHYAETDVNKRREYLTAVMTSPDSSTESQSLHLP 1017
 QY 989 EYVVPYTIHLLADPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNENSHAFIRKWNENIK 1048
 DB 1018 DYMLAFAIPVLVHDPRFTNHEDYVQLRMEKCLRFLEPLMAKRETFVHSFYKQLQLIK 1077
 QY 1049 QTKDAQCPDDAKNKKLYTCDVAMNIMSK-----STTYSLESPKDPVLPARFPTQ 1101
 DB 1078 HRFSIG-SKRONYKMWALCDLAMYIIDSFPFDGNTSTFSMPL-----ALPEMYKPE 1132
 QY 1102 D-KNFSNTKNYLPPEKSFPTPGKPTNYLGAVNKPLSSAGSQSQTKS 1149
 DB 1133 AVANFQNDVYIPLDVYT-----LGA--KTSKAAATAMTTS 1167

RESULT 6
 Q9HFF5
 ID Q9HFF5 PRELIMINARY; PRT; 1205 AA.

QY 692 FKNTGSKIEEDPHTRSALLPVLHHKSKGPPROAKYAIHCIAIFSSKETQFAQIFPL 751
DE |||||
QY 661 FKNTGSKIEEDPHTRSALLPVLHHKSKGPPROAKYAIHCIAIFSSKETQFAQIFPL 720
DE |||||
QY 752 HKSLDPSNLEHLITPLVTIIGHTALLAPDOFAAPKWSWATFTVKDLMDRLPGKKTTL 811
DE |||||
QY 721 HKSLDPSNLEHLITPLVTIIGHTALLAPDOFAAPKWSWATFTVKDLMDRLPGKKTTL 780
DE |||||
QY 812 WVPDEVPETMVKIQAIKMMVRLGLMKNHSGSTSTLRLLTTILHSDGDLTEQGRIS 871
DE |||||
QY 781 WVPDEVPETMVKIQAIKMMVRLGLMKNHSGSTSTLRLLTTILHSDGDLTEQGRIS 840
DE |||||
QY 872 KPDMSRLRLAAGSAIVKLAQPCYHEITLQYQALCALAINDECYQVQVQVFAQLKHGLS 931
DE |||||
QY 841 KPDMSRLRLAAGSAIVKLAQPCYHEITLQYQALCALAINDECYQVQVQVFAQLKHGLS 900
DE |||||
QY 932 RLRLPLEYMAICALCAKDPVRRRAHARQCLVKNINVRREYLKQHAAYSEKLLSLLPEV 991
DE |||||
QY 901 RLRLPLEYMAICALCAKDPVRRRAHARQCLVKNINVRREYLKQHAAYSEKLLSLLPEV 960
DE |||||
QY 992 VPYTHLLAHDPPYVKKVQDIEQLKDVKECLWFVLEILMAKNENSHAFIRKVENIKQTK 1051
DE |||||
QY 961 VPYTHLLAHDPPYVKKVQDIEQLKDVKECLWFVLEILMAKNENSHAFIRKVENIKQTK 1020
DE |||||
QY 1052 DAQGPDDAKMEKLYTCOVAMNIMSKSTYSLESPKDPVLPARFTQPKNFSTKNY 1111
DE |||||
QY 1021 DAQGPDDAKMEKLYTCOVAMNIMSKSTYSLESPKDPVLPARFTQPKNFSTKNY 1080
DE |||||
QY 1112 LPPEKSPFTGPKTKTNVGLAVNKLPLSAGKOSQTKSSRMETVSNASSNSPPSGRIK 1171
DE |||||
QY 1081 LPPEKSPFTGPKTKTNVGLAVNKLPLSAGKOSQTKSSRMETVSNASSNSPPSGRIK 1140
DE |||||
QY 1172 GRLDSEMDHSENYDTMSSPLPKGKSKDRDSDLVLRSELEKPRGKKTPTVTEQEKLG 1231
DE |||||
QY 1141 GRLDSEMDHSENYDTMSSPLPKGKSKDRDSDLVLRSELEKPRGKKTPTVTEQEKLG 1200
DE |||||
QY 1232 DDLTKLVOEQPKGQSRKRGHTASESDQOWPEEKRLKEDILENEDEQNSPPKKGRR 1291
DE |||||
QY 1201 DDLTKLVOEQPKGQSRKRGHTASESDQOWPEEKRLKEDILENEDEQNSPPKKGRR 1260
DE |||||
QY 1292 RPPKPLGGCTPKEPTMTKSKGSKKSGPPAPPEEEEEERQSGNTEQSKSKQHRVSR 1351
DE |||||
QY 1261 RPPKPLGGCTPKEPTMTKSKGSKKSGPPAPPEEEEEERQSGNTEQSKSKQHRVSR 1320
DE |||||
QY 1352 AQRAESPSSAISTQSTQPKGRPSKTPSPQPKNV 1391
DE |||||
QY 1321 AQRAESPSSAISTQSTQPKGRPSKTPSPQPKNV 1360
DE |||||

RESULT 4

QY4D4 PRELIMINARY; PRT; 851 AA.
AC QY4D4
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE KIAA0648 PROTEIN (FRAGMENT).
GN KIAA0648.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RF SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.*;
RU DNA Res. 5:169-176(1998).

DR EMBL; AB014548; BAA31623.1; -.
FT NON_TER 1
SQ SEQUENCE 851 AA; 94791 MW; 4F00E954B8354807 CRC64;

Query Match 37.5%; Score 2698.5; DB 4; Length 851;
Best Local Similarity 62.1%; Pred. No. 2e-156;
Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;

QY 477 LYLYATLDLNAVKALNEMWKCQNLLRHQVKDLDLIKQPKTDASVKAIFSKVMVITRL 536
DB 1 LYLYASLDLNAVKALNEMWKCQNLLRHQVKDLDLIKQPKTDASVKAIFSKVMVITRL 60
QY 537 PDPKAQDFMFKFTQVLEDDDKIRKQLEVSVPTSCSKQAEGCVREITKKGKLNPKQPTNP 596
DB 61 PDPKAQDFMFKFTQVLEDDDKIRKQLEVSVPTSCSKQAEGCVREITKKGKLNPKQPTNP 120
QY 597 FLEMKFLERLARIPIVHIDTESALIKOVNKSIDGTADDEGVPTDQAIRAGLELLKVL 656
DB 121 FLEMVFLERLARIPIVHIDTESALIKOVNKSIDGTADDEGVPTDQAIRAGLELLKVL 180
QY 657 SFTHTPISPHSAETESLILACLMDDDEKVAEALQIFKNTGSKIEEDPHIRSAIPLVHH 716
DB 181 SFTHTPISPHSAETESLILACLMDDDEKVAEALQIFKNTGSKIEEDPHIRSAIPLVHH 240
QY 717 KSKGPPROAKYAIHCIAIFSSKETQFAQIFPELHKSLDPSNLEHLITPLVTIGHIAL 776
DB 241 KAKRGTPHOAKQAVHCIAIFSSKETQFAQIFPELHKSLDPSNLEHLITPLVTIGHIAL 300
QY 777 APDOFAAPKSWATFTVKDLMDRLPGKKTTLKLVWDEEVSPEVMVKIOAKMMVRWL 836
DB 301 APDOFAAPKSWATFTVKDLMDRLPGKKTTLKLVWDEEVSPEVMVKIOAKMMVRWL 360
QY 837 LGMKNHSGSTSTLRLLTTILHSDGDLTEQKISKPDMSRLRLAAGSAIVKLAQPCYH 896
DB 361 LGMKNHSGSTSTLRLLTTILHSDGDLTEQKISKPDMSRLRLAAGSAIVKLAQPCYH 420
QY 897 EITLQVQLCALAINDECYQVQVQVFAQLKHGLSRLRLPLEYMAICALCAKDPVKERR 956
DB 421 EITLQVQLCALAINDECYQVQVQVFAQLKHGLSRLRLPLEYMAICALCAKDPVKERR 480
QY 957 HARCLVKNINVRREYLKQHAAYSEKLLSLLPEYVVPYTIHLLAHDPPYVKKVQDTEQLK 1016
DB 481 HARCLVKNINVRREYLKQHAAYSEKLLSLLPEYVVPYTIHLLAHDPPYVKKVQDTEQLK 540
QY 1017 VKECLWFVLEILMAKNENSHAFIRKVENIKQTKDAQGPDDAKMEKLYTCOVAMNII 1076
DB 541 IKECLWFVLEILMAKNENSHAFIRKVENIKQTKDAQGPDDAKMEKLYTCOVAMNII 600
QY 1077 MSKSTYSLESPKDPVLPARFTQPKNFSTKNYLPPEKMSFTTQPKKPTNVLGAVNK 1136
DB 601 NSKALCNADSPKDPVLPARFTQPKNFSTKNYLPPEKMSFTTQPKKPTNVLGAVNK 660
QY 1137 PLSSAGQSKQTKSSRMETVS--NASSSNPSPPGRIKGR-LDSSEMDHSENE--YTWS 1191
DB 661 PLSSAGQSKQTKSSRMETVS--NASSSNPSPPGRIKGR-LDSSEMDHSENE--YTWS 720
QY 1192 PLPGKSKDRDSDLVLRSELEKPRGKKTPTVTEQEKLGMDLTKLVOEQPKGQSRK 1251
DB 721 VTPVKNID-----PVKNKE--INSQAT-----QGNISDRGCK 752
QY 1252 RGHATASEDE--QOWPEEKRLKEDILENEDEQNSP--PKGKRGRRPKPLG--GTPKEEPTM 1308
DB 753 RTVTAAGAEINQOKTDEK-----VDESGPPAPSPRRGRPKSESGQGNATKNDN 803
QY 1309 KTSKKGSKKSGPPAPPEEEEEERQSGNTE 1338
DB 804 KPINKGRAR----AAVGQSPGGLEAGNAK 829

RESULT 5
QYV646 PRELIMINARY; PRT; 1218 AA.
ID QYV646


```
QY 421 LQSAAGKDAKQIAWIKDKLLHIYYQNSIDRLLVERIFAQYVMPHNLETTERMKCLYYL 480
Db 421 LQSAAGKDAKQIAWIKDKLLHIYYQNSIDRLLVERIFAQYVMPHNLETTERMKCLYYL 480
QY 481 YATLDLNAVKALEMMKQNLRLHQVKDLDLIKQPKTASVKAIFSKVWVITRNLPDPG 540
Db 481 YATLDLNAVKALEMMKQNLRLHQVKDLDLIKQPKTASVKAIFSKVWVITRNLPDPG 540
QY 541 KAQDMFKFTQVLEDEKTRKOLEVLVSPSCSKQAEGCVREITKLGPNKOPTNPFLEM 600
Db 541 KAQDMFKFTQVLEDEKTRKOLEVLVSPSCSKQAEGCVREITKLGPNKOPTNPFLEM 600
QY 601 IKFLLERIAVPHIDTESISALIKQVKNKSIDGTADDEGVPTDQAIRAGLELLKVLFSFH 660
Db 601 IKFLLERIAVPHIDTESISALIKQVKNKSIDGTADDEGVPTDQAIRAGLELLKVLFSFH 660
QY 661 PISFSAETFESLACLMKDDKVAEALQIFKNTGSKITEEDPPIHRSALLPVLHHSKK 720
Db 661 PISFSAETFESLACLMKDDKVAEALQIFKNTGSKITEEDPPIHRSALLPVLHHSKK 720
QY 721 GPPROAKYAIHCITAFSSKETQFAQIFPLHKSIDPSNLEHLITPLVTIGHIALLAPDQ 780
Db 721 GPPROAKYAIHCITAFSSKETQFAQIFPLHKSIDPSNLEHLITPLVTIGHIALLAPDQ 780
QY 781 FAAPKMSWATFIVKDLLMNDRLPGKTTKLWVPDEVSPETMVKIQAIKMMVWLLGMK 840
Db 781 FAAPKMSWATFIVKDLLMNDRLPGKTTKLWVPDEVSPETMVKIQAIKMMVWLLGMK 840
QY 841 NNHSGSTLRLTLTILHSDGDLTPQGIKSKPDMRSLRAGASIVKLAQEPICYHEIT 900
Db 841 NNHSGSTLRLTLTILHSDGDLTPQGIKSKPDMRSLRAGASIVKLAQEPICYHEIT 900
QY 901 LEQYOLCALAINDECYQVQVQFAQLHGLSLRLPLEYMAICALCAKDPVKERRAHQ 960
Db 901 LEQYOLCALAINDECYQVQVQFAQLHGLSLRLPLEYMAICALCAKDPVKERRAHQ 960
QY 961 CLVKNNVRREYLKQHAAYSEKLLSLLPEYVVPYTHLLAHDPDYVVKQDIEQLKDVKEC 1020
Db 961 CLVKNNVRREYLKQHAAYSEKLLSLLPEYVVPYTHLLAHDPDYVVKQDIEQLKDVKEC 1020
QY 1021 LWFVLEILMAKNNNSHAFIRKMWENIKOTKDAQGPDADAKMEKLYTVCDAVNNIIMSKS 1080
Db 1021 LWFVLEILMAKNNNSHAFIRKMWENIKOTKDAQGPDADAKMEKLYTVCDAVNNIIMSKS 1080
QY 1081 TTYLESPPKDPVLPARFFTPQDKNFNTKNYLPPEMKSFPTPKPKPTTIVLGAVNKPSS 1140
Db 1081 TTYLESPPKDPVLPARFFTPQDKNFNTKNYLPPEMKSFPTPKPKPTTIVLGAVNKPSS 1140
QY 1141 AGKQSQTKSRMETVSNASSSSNPSPGRIKGRDSEMDHSENEIDYTWSSPLPGKSKSD 1200
Db 1141 AGKQSQTKSRMETVSNASSSSNPSPGRIKGRDSEMDHSENEIDYTWSSPLPGKSKSD 1200
QY 1201 RQSDLVRLSELEKPRGRKKTPTVTEQEKILGMDDLTKLVQEQPKGQSRKRGHTASESD 1260
Db 1201 RQSDLVRLSELEKPRGRKKTPTVTEQEKILGMDDLTKLVQEQPKGQSRKRGHTASESD 1260
QY 1261 EQOWPEEKLEKEDILENEDEQNSPPKKGRRPPKPLGGGTKEEPTMTKSKGSKKSG 1320
Db 1261 EQOWPEEKLEKEDILENEDEQNSPPKKGRRPPKPLGGGTKEEPTMTKSKGSKKSG 1320
QY 1321 PPAPEEEEEEQSGNTEQKSKSHQVRSRAQQAQRAESPESSESAIESTQSTPOKGRGRPSK 1380
Db 1321 PPAPEEEEEEQSGNTEQKSKSHQVRSRAQQAQRAESPESSESAIESTQSTPOKGRGRPSK 1380
QY 1381 TPSPSQPKNV 1391
Db 1381 TPSPSQPKNV 1391
```

RESULT 3

Q9Y215

ID Q9Y215 PRELIMINARY; PRT: 1416 AA.

AC Q9Y215;

```
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE KIAA0979 PROTEIN (FRAGMENT).
GN KIAA0979.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL: AB023196; BAA76823.1; -.
DR InterPro: IPR000637; -.
DR PRINTS: PR00929; ATHOOK.
FT NON_TER 1
SQ SEQUENCE 1416 AA; 161192 MW; 0BC682D7F4657424 CRC64;
```

```
Query Match 97.2%; Score 6994; DB 4; Length 1416;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1356; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 32 VRLKMWVKTFTMDMDQDSEEEKELYNLALHLASDFELKHPGKDVRLLVACCLADIFRIY 91
Db 1 VRLKMWVKTFTMDMDQDSEEEKELYNLALHLASDFELKHPGKDVRLLVACCLADIFRIY 60
QY 92 APEAPYTPDKLKIDIPMTITROLKGLDTSKPOFNRYFYLLLENIAVWYSNYICFELEDSN 151
Db 61 APEAPYTPDKLKIDIPMTITROLKGLDTSKPOFNRYFYLLLENIAVWYSNYICFELEDSN 120
QY 152 EIFTQLYRTLFSVINGNHQKVHMHVMDLMSIICSGDVTSQELLOTVLVNLVPAHKNLN 211
Db 121 EIFTQLYRTLFSVINGNHQKVHMHVMDLMSIICSGDVTSQELLOTVLVNLVPAHKNLN 180
QY 212 KQAYDLAKALLKRTAQAIPEYITTFNFQVLMGKTSISDLSEHFVLDLILELYNIDSHLL 271
Db 181 KQAYDLAKALLKRTAQAIPEYITTFNFQVLMGKTSISDLSEHFVLDLILELYNIDSHLL 240
QY 272 SVLPQLEFKLSNDNEERLQVYKLLAKMFGAKDSELSQNKPLWCYLGFRNDIHPRL 331
Db 241 SVLPQLEFKLSNDNEERLQVYKLLAKMFGAKDSELSQNKPLWCYLGFRNDIHPRL 300
QY 332 ECVKFASHCLMHPDLAKDLTEYLKVRSHDPEAIRHDVIVSVITAAKDKILLVNDHLLN 391
Db 301 ECVKFASHCLMHPDLAKDLTEYLKVRSHDPEAIRHDVIVSVITAAKDKILLVNDHLLN 360
QY 392 FVRERTLDKRWVRKEMGLAQIYKKYALQSAAGDAKQIAWIKDKLLHIYYQNSIDD 451
Db 361 FVRERTLDKRWVRKEMGLAQIYKKYALQSAAGDAKQIAWIKDKLLHIYYQNSIDD 420
QY 452 RLIVERIFAQYVMPHNLETTERMCKLYLYATLDLNAVKALEMMKQNLRLHQVKDL 511
Db 421 RLIVERIFAQYVMPHNLETTERMCKLYLYATLDLNAVKALEMMKQNLRLHQVKDL 480
QY 512 LKQPKTASVKAIFSKVWVITRNLPDPGKAQDFMKFTQVLEDEKIRKQELVLSPTC 571
Db 481 LKQPKTASVKAIFSKVWVITRNLPDPGKAQDFMKFTQVLEDEKIRKQELVLSPTC 540
QY 572 SKQAEGCVREITTKLGNPKOPTNPFLEMIKFLERIAVPHIDTESISALIKQVKNKSIDG 631
Db 541 SKQAEGCVREITTKLGNPKOPTNPFLEMIKFLERIAVPHIDTESISALIKQVKNKSIDG 600
QY 632 TADDEGVPTDQAIRAGLELLKVLFSFTHPISFSAETFESLACLMKDDKVAEALQI 691
Db 601 TADDEGVPTDQAIRAGLELLKVLFSFTHPISFSAETFESLACLMKDDKVAEALQI 660
```


241 LMLGTSISDLSEHFVDLILELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
301 GAKDSELASONKPLQWCVYLGRENTHVPIRLCEVAFASHCLMNHDPDLAKDTEYLKVRSH 360
301 GAKDSELASONKPLQWCVYLGRENTHVPIRLCEVAFASHCLMNHDPDLAKDTEYLKVRSH 360
361 DPEEAIHVDIVSVITAAKODILLVNDHLLNFVRETLDKRWRVRKEAMGLAQIYKKYA 420
361 DPEEAIHVDIVSVITAAKODILLVNDHLLNFVRETLDKRWRVRKEAMGLAQIYKKYA 420
421 LOSAAGKDAAKOIAWKOKLLHIYQNSIDDRLLVERIFAQVWPHNLETTERMKCLYL 480
421 LOSAAGKDAAKOIAWKOKLLHIYQNSIDDRLLVERIFAQVWPHNLETTERMKCLYL 480
481 YATDLNAVKALNMWQKONLLRHQVOKLLDLIKOPKTDASVKAIFSKVMVITRNLPPDG 540
481 YATDLNAVKALNMWQKONLLRHQVOKLLDLIKOPKTDASVKAIFSKVMVITRNLPPDG 540
541 KAQDPMKKFTQVLEDEKIRKOLEVLSVPTSCCKQAEGCVREITKGLGNPKQPTNPFL 600
541 KAQDPMKKFTQVLEDEKIRKOLEVLSVPTSCCKQAEGCVREITKGLGNPKQPTNPFL 600
601 IKFLLERAPVHIDPESISALIKOVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSETH 660
601 IKFLLERAPVHIDPESISALIKOVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSETH 660
661 PISFHSATFESLLACLKMDDEKVAEALQIPKNTGSKIEEDFPHRSALLPVLHKKSK 720
661 PISFHSATFESLLACLKMDDEKVAEALQIPKNTGSKIEEDFPHRSALLPVLHKKSK 720
721 GPPROAKYAIHICHAIFSKETQFAQIFPEPLHKS LDPNSLEHLITPLVTIGHIALLAPDQ 780
721 GPPROAKYAIHICHAIFSKETQFAQIFPEPLHKS LDPNSLEHLITPLVTIGHIALLAPDQ 780
781 FAAPKSWVATFIVKDLLMNDRLPCKTKTLWVPEEVSPETMVKIOAKMVRLLGNK 840
781 FAAPKSWVATFIVKDLLMNDRLPCKTKTLWVPEEVSPETMVKIOAKMVRLLGNK 840
841 NNHKSSTLRLTLTILHSDGLTEQGIKSKPDMSRLRLAAGSAIVKLAOPCYHEIIT 900
841 NNHKSSTLRLTLTILHSDGLTEQGIKSKPDMSRLRLAAGSAIVKLAOPCYHEIIT 900
901 LEQYOLCALAINDECYQVQVPAQKHLKGLSRLRPLEYMAICALCAKDPKERRAHARQ 960
901 LEQYOLCALAINDECYQVQVPAQKHLKGLSRLRPLEYMAICALCAKDPKERRAHARQ 960
961 CLVKNINVRREYLKQAAVSEKLLSLPEYVVPYTIHLLAHDPPVYKQDIEQLKDVREC 1020
961 CLVKNINVRREYLKQAAVSEKLLSLPEYVVPYTIHLLAHDPPVYKQDIEQLKDVREC 1020
1021 LWFVLEILMAKNENSHAFIRKAVENIKQTKDAQGPDPAKMKNEKLYTVCDAVMNIIMSKS 1080
1021 LWFVLEILMAKNENSHAFIRKAVENIKQTKDAQGPDPAKMKNEKLYTVCDAVMNIIMSKS 1080
1081 TTYLESKDPVLPARFTQDPKNSFNKNYLPPEMKSFPTGPKPTTNVGAUNKPLSS 1140
1081 TTYLESKDPVLPARFTQDPKNSFNKNYLPPEMKSFPTGPKPTTNVGAUNKPLSS 1140
1141 AKQSOTKSSRMETVSNASSSNPSSPGRIKRLDSSSEMDHSENEDEYTWSSPLPKKSDK 1200
1141 AKQSOTKSSRMETVSNASSSNPSSPGRIKRLDSSSEMDHSENEDEYTWSSPLPKKSDK 1200
1201 RDDSILVSELEKPRGRKKTPTVEQEEKLGMDLTKLVQEQPKGSQRKRGRHGTASESD 1260
1201 RDDSILVSELEKPRGRKKTPTVEQEEKLGMDLTKLVQEQPKGSQRKRGRHGTASESD 1260
1261 EQQWPEERKLKEDILENEDEQNSPPKGRGRPPKPLGGGTPKEEPTMTKSKGSKKSG 1320
1261 EQQWPEERKLKEDILENEDEQNSPPKGRGRPPKPLGGGTPKEEPTMTKSKGSKKSG 1320
1321 PPAPPEEEEEERQSQNTQKSKSKOHRVSRRAQRAESPESIAESTOSTPQKGRGRPSK 1380

Db 1321 PPAPPEEEEEERQSQNTQKSKSKOHRVSRRAQRAESPESIAESTOSTPQKGRGRPSK 1380
QY 1381 TPSPSQPKNV 1391
Db 1381 TPSPSQPKNV 1391
RESULT 2
QY451 PRELIMINARY; PRT: 1391 AA.
AC QY451.
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN.
DS AS3.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;
RT "Androgen-induced proliferative shutoff in prostate cancer cells";
RL Proc. Am. Assoc. Cancer Res. 37:223-223(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=92229875; PubMed=10215036;
RA Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;
RT "Early gene expression during androgen-induced inhibition of
RT proliferation of prostate cancer cells: a new suppressor candidate on
RT chromosome 13, in the BRCA2-Rb1 locus";
RL J. Steroid Biochem. Mol. Biol. 68:41-45(1999).
DR EMBL; U95825; AAD22134.2; -;
DR InterPro; IPR002687; -;
DR ProDom; PD004104; -; 1.
SQ SEQUENCE 1391 AA; 158035 MW; F58AEEB5AD6D9479 CRC64;

Query Match 98.7%; Score 7101; DB 4; Length 1391;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1377; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
QY 1 MAHSTRNDGKITVPGVKEISKISKEEMVRLKMWVKTMDMDQDSEEEKELYLNLA 60
Db 1 MAHSTRNDGKITVPGVKEISKISKEEMVRLKMWVKTMDMDQDSEEEKELYLNLA 60
QY 61 LHLASDFELKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLDFIMFITRLQKLEDT 120
Db 61 LHLASDFELKHPGKDVRLVACCLADIFRIYAPEAPYTPSPDKLDFIMFITRLQKLEDT 120
QY 121 KSPQNRIFYLLENATWVKNYCNICELEDSNEIFQLYRTLFESVINGHKNQVHHMVDL 180
Db 121 KSPQNRIFYLLENATWVKNYCNICELEDSNEIFQLYRTLFESVINGHKNQVHHMVDL 180
QY 181 MSSICEGDTVSQELLDVTLVNLVPAHNLNKOAYDLAKALKRTAQAEPIYITFFNOV 240
Db 181 MSSICEGDTVSQELLDVTLVNLVPAHNLNKOAYDLAKALKRTAQAEPIYITFFNOV 240
QY 241 LMLGTSISDLSEHFVDLILELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
Db 241 LMLGTSISDLSEHFVDLILELYNIDSHLLSVLPQLEFKLSNDNEERLQVVKLLAKMF 300
QY 301 GAKDSELASONKPLQWCVYLGRENTHVPIRLCEVAFASHCLMNHDPDLAKDTEYLKVRSH 360
Db 301 GAKDSELASONKPLQWCVYLGRENTHVPIRLCEVAFASHCLMNHDPDLAKDTEYLKVRSH 360
QY 361 DPEEAIHVDIVSVITAAKODILLVNDHLLNFVRETLDKRWRVRKEAMGLAQIYKKYA 420
Db 361 DPEEAIHVDIVSVITAAKODILLVNDHLLNFVRETLDKRWRVRKEAMGLAQIYKKYA 420

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:02:24 : Search time 39.87 seconds
(without alignments)
4615.909 Million cell updates/sec

Title: US-09-512-581-2
Perfect score: 7193
Sequence: 1 MAHSKTRTNDGKITYPGCVK.....QKGRGRSPKTPSPSQPKNV 1391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------|--------------------|
| 1 | 7155 | 99.5 | 1447 | Q9NTI5 | Q9nti5 homo sapien |
| 2 | 7101 | 98.7 | 1391 | Q9V451 | Q9v451 homo sapien |
| 3 | 6994 | 97.2 | 1416 | Q9Y2I5 | Q9y2i5 homo sapien |
| 4 | 2698.5 | 37.5 | 851 | Q9Y4D4 | Q9y4d4 homo sapien |
| 5 | 1982.5 | 27.6 | 1218 | Q9V646 | Q9v646 drosophila |
| 6 | 833.5 | 11.6 | 1205 | Q9HFF5 | Q9hff5 schizosacch |
| 7 | 808 | 11.2 | 1638 | Q9FILO | Q9fil0 arabidopsis |
| 8 | 807 | 11.2 | 1579 | Q9XTF6 | Q9xtf6 caenorhabdi |
| 9 | 747.5 | 10.4 | 1596 | Q9UVY6 | Q9uyv6 sordaria na |
| 10 | 656 | 9.1 | 125 | Q9H5N8 | Q9h5n8 homo sapien |
| 11 | 632.5 | 8.8 | 1506 | Q94076 | Q94076 emericella |
| 12 | 309.5 | 4.3 | 390 | Q94237 | Q94237 schizosacch |
| 13 | 303 | 4.2 | 780 | Q9SAI0 | Q9sai0 arabidopsis |
| 14 | 264 | 3.7 | 990 | Q9S9P0 | Q9s9p0 arabidopsis |
| 15 | 248.5 | 3.5 | 2158 | Q9Y6Y3 | Q9y6y3 homo sapien |
| 16 | 242.5 | 3.4 | 2265 | Q9Y6Y4 | Q9y6y4 homo sapien |
| 17 | 234.5 | 3.3 | 852 | Q9SZ55 | Q9sz55 arabidopsis |
| 18 | 231.5 | 3.2 | 2253 | P70012 | P70012 xenopus lae |
| 19 | 230 | 3.2 | 2501 | Q9NCW7 | Q9ncw7 drosophila |

ALIGNMENTS

```

RESULT 1
Q9NTI5 ID Q9NTI5 PRELIMINARY; PRT; 1447 AA.
AC Q9NTI5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 164.7 KDA PROTEIN.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rhodes S., Huckle E.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137201; CAB69911.1; -
KW Hypothetical protein.
SQ SEQUENCE 1447 AA; 164666 MW; 145C30308EA3EFD5 CRC64;

```

```

Query Match 99.5%; Score 7155; DB 4; Length 1447;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1387; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAHSKTRTNDGKITYPGCVKEISDKISKEEMVRRLKMKVKTFFMDMDODSEEEKELYLNLA 60
DB 1 MAHSKTRTNDGKITYPGCVKEISDKISKEEMVRRLKMKVKTFFMDMDODSEEEKELYLNLA 60

QY 61 LHLSADFFLKHGPKDVRLLVACCLADIFRIYAPEAPYTPSPDKLKDIFMFTITROLKGLDT 120
DB 61 LHLSADFFLKHGPKDVRLLVACCLADIFRIYAPEAPYTPSPDKLKDIFMFTITROLKGLDT 120

QY 121 KSPQFNRYFYLLENIAWKSYNICFELEDSNEIFTQLYRTLFVSNNHGNKQVHHMVDL 180
DB 121 KSPQFNRYFYLLENIAWKSYNICFELEDSNEIFTQLYRTLFVSNNHGNKQVHHMVDL 180

QY 181 MSSIIICGDTVSGLDVTLVNLPVPAHKNLKNQAYDLAKALLKRTAQAIPEYITTFNQV 240
DB 181 MSSIIICGDTVSGLDVTLVNLPVPAHKNLKNQAYDLAKALLKRTAQAIPEYITTFNQV 240

QY 241 LMLGKTSISLSEHFVDLILELYNIDSHLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300

```

Q9jmx8 helicobacte
Q9i7u4 drosophila
Q94248 schizosacch
Q07569 entamoeba h
Q92lv0 helicobacte
Q080784 arabidopsis
Q9nfs3 drosophila
Q25262 helicobacte
Q9uik5 schizosacch
Q07380 saccharomyc
Q9uql4 homo sapien
Q9ujs7 homo sapien
Q9xze3 amoeba prot
O15019 homo sapien
O35482 rattus norv
Q9nev4 caenorhabdi
Q18033 caenorhabdi
Q9v628 drosophila
O14617 homo sapien
Q9w3d1 drosophila
P91257 caenorhabdi
Q9v507 drosophila
Q99968 homo sapien
Q17362 caenorhabdi
O01761 caenorhabdi
Q26216 plasmodium

THIS PAGE BLANK (USPTO)

```

1054 .....GlnGlyPro.....AspAspAlaLysMetAsnGlu 1063
1809 TCCTTCAGCAGCAAGGAGGAGGCTCTAGAGCAGCGGAAGAGGCTGCAG 1858
1064 LysLeuTyrThrValCysAspValAlaMetAsnIle.....IleMetSe 1078
1859 GCAGACTGTCAGCTCTGCTGGAGAAATGCAGCTTCAAGTCCAGTC 1908
1078 rlySerThrThrTyrSerLeuGluSer..... 1087
1909 CCCTTCATAAAATGAGCCTTGAGCTACCGGAGAGCAGCTATGTAGCA 1958
1088 .....ProLysAspProValLeuProAlaArgPhePheThrGlnProAsp 1102
1959 TGGTCCCGCAGCAATGAGAAGCTGTCGAAGACTTATGATACCACTGTGAT 2008
1103 LysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSer... 1118
2009 GACAATGAGAGCTATTACAAGAGTTACACGACCACCCAGAGCAGCA 2058
1119 .PhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValA 1135
2059 GAGCTTTCTCAAGACTATGACAGCAGCAGCAGCTCCAGCTGAGGCTATG 2108
1135 snLysProLeuSerSerAlaGly.....LysGlnSer 1145
2109 GGAAGAGTTACTCCACTACCAGCAACAGCAGCATTTACCTATATAAGAAGCT 2158
1146 GlnThrLysSerSerArgMetcluThrValSerAsnAlaSerSerSe 1162
2159 TAGCGCAGCACCAGTAGCTCTGACACCTGCCAGAGAGTTTGTTCAGCAG 2208
1162 r.....AsnProSerSerPro..... 1167
2209 CTGCACTGACGAGGAACCTGCTGACGCTCAAGACATGGAGCCCTTGAGG 2258
1168 .....GlyArgIleLysGlyArgLeu 1174
2259 AAATGGTTGTGAAAGTGCTGATCAAGCTGCAGCGGTGCAGGCCATGTAC 2308
1175 AspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSe 1191
2309 CAGATAAGCCAGGAGGACACAGCAGCTGCAAGAG.....CAGATGGA 2352
1191 rProLeuProGlyLysLysSerAspLysArgAspSerAspLeuValA 1208
2353 AAAGTTACTGGCCACAGCAGAAAGACCTGAAGGAGAGCTGGATGCTGTG 2402
1208 rSerGlu.....LeuGluLysProArgGly... 1216
2403 AAAGGGAGTTCAAGGAGTGCAATGCGTGTGAAGGCCCATGGCCCCC 2452
1217 ...ArgLysLysThrProValThrGlnGlnGluLysLeu..... 1229
2453 CAGAAGCACAAGATGAGATCAAAAGTAAGTGCACCAAGCTCGGGAGCT 2502
1230 .....GlyMetAspLeuThrLysLeu.....V 1238
2503 GCAGCTGCAATACAGGCTAGCATGGATGATGAGCAGCGGGCGGCTTCTGGTAG 2552
1238 aGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHis 1254
2553 TGCAGGAGCAG.....CTGAGGGGCGAG 2575
1255 ThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArg.... 1269
2576 CTGCAAGTGTCCAGGAGGAGCTCCCGCAGCTCAGGGAGAGAGGCGCTTC 2625
1270 ...LeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProp 1285
2626 TGTTCGTAAGAGAGCCCGGGGGAAGAACTCTAATAAGAACATGAACAAGA 2675
1285 rOLysLysGly...LysArgGlyArgProProLysProLeuGlyGlyGly 1300
2676 ATGCCAATGGGTTAAATGAAAAAGTGACCAAGCCATCGTCGGATACT 2725
1301 ThrProLysGluGluProThrMetLysThrSerLys..... 1312
2726 TCTGAGAGCGACCTTGAGACCAGAAAGAGTCTGGAGGTAGTCTGTACTA 2775
1313 .LysGlySerLysLysSerGlyProProAlaProGluGluGluGluG 1329
2776 CAAGCCAGCCAGAGAGAAATTTAGATGGACTAGCAAAAGAGAGGAGAAAGA 2825
1329 LuGluGlu..... 1331
2826 AAAGAGAGATGGAGGAGGAAAAAGAAAGTGAAGAGGAAAGCAAGGAG 2875
1332 .....ArgGlnSerGln 1335
2876 CAGTGTGGGATGAGCTAGTTGCTGAGCCAGCAGATCTCGAGAAAGCTAA 2925
1335 yAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgA 1352
2926 ATCCACAGAGATCAGGAGGAGAAATGAAGAGGACAAAGAGAGAGGAGA 2975
1352 laGlnGlnArgAlaGluSerProGluSerSerSerAlaIleGluSerThrGln 1368
2976 AGGAAGAAGACAGTGAAGAGAGGAGGATGACGCGGACTCTTCCCTTGA 3025
1369 SerThrProGlnLysGlyArgGlyArgProSerLys.....Th 1381
3026 ATCCCGGAAGAAATAACCCCTCAGACTTCCGAGAGCAAAAGTCATC 3075
1381 rProSerProSerGlnPro 1387
3076 CCTACCCCAATCCCCC 3094

```



```

556 GACTTCAGAAATGGATTCTTAGAGCTGATCCTGAAATG..... 595
      :|||
499 InAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
      :|||
596 ..CAGTGTTCAGCGACGACGCTACGGGAT..... 622
      :|||
516 ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIle 532
      :|||
623 .....GCTGACAGCAGATGCATGCATGAAGACAAAGTGTCCAGGAAT 666
      :|||
532 eThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysP 549
      :|||
667 GTGTGTGTGAGTGTGAAGAG.....CTACAGCATC 695
      :|||
549 heThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluVal 565
      :|||
696 ATGCCAGAGGTCACTGAGGAGGACAGAGG...CGGCTGCACAGAGCTC 742
      :|||
566 LeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgL 582
      :|||
743 .....AAGTGTGCTCAGAAAT..... 757
      :|||
582 uIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuG 599
      :|||
758 ..... 758
599 luMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
      :|||
759 AGGTGCTTCGGTTTCAGACCTCCACAGTGC.....ACCCAGAAAT 799
      :|||
616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyTh 632
      :|||
800 GAGGAGCTGAAGTCCAGACTCTGTACCTGCAGAAAAAATATACTAG 849
      :|||
632 rAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaG 649
      :|||
850 CCAGGATGAGCAGAACGAGCTCTTGAAGATGCAGCTGCAACTTCAGACTG 899
      :|||
649 lyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHis 665
      :|||
900 AGTCCGGAGCTCAAGTCAATGATGAAATCCACA...CTGTGAGAAAAACCAG 946
      :|||
666 SerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspG 682
      :|||
947 AGTGAGAGAGAGTACTGTCCCGGCTGCAGAACTGCACCTCCAGCACCA 996
      :|||
682 uLysValAla..... 685
997 GAACGTACATGTGAGAAAGGAAAGCTGTGGAACGACGACGAGCTGC 1046
      :|||
686 .....GluAlaAlaLeuGlnIlePheLysAsnThr 695
1047 AGGAGGAGCTGCAGTGCCATGAGGCGAGCTGCAGCACCTCAGGGATACG 1096
      :|||
696 GlySerLysIleGluGluAsp.....PheProHisIleArg 707
      :|||
1097 GTGGCTCTCTTCAAGAGAGCAATGAGAGGACACAGAGCGACCGCTCA 1146
      :|||
707 gSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProA 724
      :|||
1147 GCTTCAGGAGATGAAGCAGCTGTACCAGGCCACGAGGAGGAGCTGGAGC 1196
      :|||
724 rgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
      :|||
1197 GGCAG..... 1201
      :|||
741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPr 757
      :|||
1202 .....AAGCACATGTATGACCACTGGAGCAGGACCTCTCTCGT 1239
      :|||
757 oSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleA 774
      :|||
1240 CTCCAGCTGGAG..... 1252
774 laLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAla 790
1252 ..... 1252
791 ThrPheIleValLysAspLeuLeuMetAsnAspArgLeuPro..... 804
      :|||
1253 .....CTGAAGAGCTCAAGGCTCCACCCCATTCGCGAGGACAA 1293
      :|||
805 GlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProG 821
      :|||
1294 AGGAAAGTGTCTAATAAG..... 1312
821 luThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeu 837
1312 ..... 1312
838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLe 854
1313 .....TG 1314
854 uThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIle 871
      :|||
1315 TGACACACTGCTCTCCAGACTGACAGAAATTCAGAGAAAGTACAGAGCCA 1364
871 erLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVal 887
      :|||
1365 GCCAGAGGAGATGGGGCAGCTG..... 1387
888 LysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTy 904
1388 .....CAGAT 1392
904 rgLnuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnV 921
      :|||
1393 GGAGCAGCTGTGAGCTCTCGGAGGATCAGAGGAGTGCAGGAGGAGCAG. 1441
921 alPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
      :|||
1442 .....GCCAGCTGCAGGAAGAGCTGCACAGGCTCACACTGCCACTG 1483
938 GluTyrMetAlaIleCysAla..... 944
1484 CCAAGAGTGGCCTCTTACTCAAGAGTCAGAGCTACTCACCAGTTAGA 1533
945 ....LeuCysAlaLysAspProValLysGluArgAlaHisAlaArgG 960
      :|||
1534 AGACCTGTGTGAGCTGCAGCTGCTCTACCAAGGCATGCAGGAGGACAGA 1583
960 lncCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
      :|||
1584 AGAAGCTGATACAGAACCAAGACTGT.....GTATTAAAGAACAA 1624
977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPr 993
1625 TTAGAGATCCACCAAGAGCTGCAGCTTTCAAGAG..... 1660
993 oTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnA 1010
      :|||
1661 .....TCTCATTTCCAGGAAGTGTGTGGAGAATCCCG 1691
1010 spIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGlu 1026
      :|||
1692 ATGATTCCAAATTTGGTAAAGTCTCTCCAAATGT..... 1723
1027 lIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMe 1043
      :|||
1724 .....AATCGAAACAAACAATCCAAGCTGCTCATGGAGCA 1758
1043 tValGluAsnIleLysGlnThrLysAspAla..... 1053
      :|||
1759 GATGCAGCCCTGCAGGTGATGTATGACCGCGGTTCAGGACGAGCAGGAGC 1808
```



```
677 uLysMetAspGluTyrValAlaGluAlaLeuGlnIlePheLysA 694
|||||:|||||
715 .AGGTAGAGATGAAAAA..... 732
694 snThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaLeu 710
|||||:|||||
733 ..GAAGGAGTGGAGTGGAGACCAATCAAAAGTGGAGAGCAAA... 777
711 LeuProValLeuHisLysSerLysLysGlyProArgGlnAlaLy 727
777 ..... 777
727 sTyRAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnP 744
|||||:|||||
778 .....GATGCAAGGAAGATGAG. 795
744 heAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
795 ..... 795
761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuAl 777
795 ..... 795
777 aProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleV 794
795 ..... 795
794 aLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLys 810
|||||:|||||
796 ..AAAGAGGAGACAAATGATGATAAAGAAAGATGAAAAAGAGAGCAAG 843
811 LeuTrpValProAspGluGluValSerProGluThrMetValLysIleG 827
843 ..... 843
827 nAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHis 844
|||||:|||||
844 .....GGTCTTAAA..... 852
844 erLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
|||||:|||||
853 ..AAGCGTGGAAAGGACG.....AGTTCT 876
861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerAr 877
|||||:|||||
877 GGAGGAAGGTTCGCGAGAGAAATAAACCCAGGAA..... 912
877 gLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProC 894
|||||:|||||
913 .....GTAAAAAGGATGCGAGCGCT. 933
894 yStyHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAla 910
933 ..... 933
911 IleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln..... 924
|||||:|||||
934 .....AGGACCTCTTCTCTGATCGCCCTGT 959
925 .LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMeta 941
|||||:|||||
960 CGGTGACCGGAATCTGTGAGGCTT..... 987
941 latLeCysAlaLeuCysAlaLysAspProValLysGluArgAlaHis 957
|||||:|||||
988 .....GTTGCATTGATGATAAAGACTCTCTCGAAAGAAATTCGTTGTGAA 1032
958 ...AlaArgGlnCysLeuValLysAsnIle...AsnValArgArgGluTy 972
|||||:|||||
1033 AAGGGCGGAGTGCATATCTCAAGATATTTCCTCAATGTTCTGCTAACAAAGGT 1082

972 rLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProG 989
|||||:|||||
1083 AATGAGGAAGAGG.....TCTGATGAACCTTTGAACCTGCTT.... 1119
989 luTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyr 1005
|||||:|||||
1120 .....CACCAATTCTATTGCTGGGAGGAGA 1146
1006 valLysValGlnAspIleGlu...GlnLeuLysAspValLysGluCysLe 1021
|||||:|||||
1147 GGAAGGCTGCTCAGATCAAGACAAACATATTGGGCTTCTCTGTTTCGT 1196
1021 uTrpPheValLeuGluIleLeuMetalLysAsnGluAsnAsnSerHisA 1038
|||||:|||||
1197 TTGG.....CATG 1204
1038 laPheIleArgLysMetValGluAsnIleLysGln..... 1049
|||||:|||||
1205 GAGATGAGAAAGAAAGCAAAAGAAAGTAAAGAAAGCTTGAGAAATGC 1254
1050 ThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTy 1066
|||||:|||||
1255 ACCAAA.....GAGAAACTGTG 1271
1066 rThrValCysAspValAlaMetAsnIleMetSerLysSerThrThrT 1083
|||||:|||||
1272 GGAGTTTGTGATGTG...TTGGACATACACATTACCAAGGCTACAACA. 1317
1083 yrSerLeuGluSerProLysAspProValLeuProAlaArgPheThr 1099
|||||:|||||
1318 .....AAGAAGGAAGATATTATT..... 1335
1100 GlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMe 1116
1336 .....AC 1337
1116 tLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyA 1133
|||||:|||||
1338 AAAACTGTTTGTGAGTTTGGAGAAACCT.....CATGTGACAGGTG 1378
1133 laValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys... 1148
|||||:|||||
1379 ATGTGACCGGTGACACTACACTTCTGAGAAAGAGAAAGTCAAGTAAGGGA 1428
1149 SerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSe 1165
|||||:|||||
1429 GCAAAAGCGAAGAGACTCCCAAGAAACTTCACCTACAGCTGGGAGTTC 1478
1165 rSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHis 1182
|||||:|||||
1479 ATCTCTCCAAACGA.....TCAGCAAGAGCCCAA 1507
1182 erGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSer 1198
|||||:|||||
1508 AAAAGTCTGAAGAGCAACAAAAGTTGTC.....AAAAACAGT 1545
1199 AspLysArgAspSerAspLeuValArgSerGluLeuGluLysProAr 1215
|||||:|||||
1546 TTAGCTCATTTCTGATGATAA.....TCTGAAGAAAGAAAGAAAGA 1586
1215 gGlyArgLysLysThrProValThrGluGlnGluLysLeuGlyMeta 1232
|||||:|||||
1587 AGAGAAAAA.....CAAGAGGAGGAGAAG..... 1611
1232 spAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnArg 1248
|||||:|||||
1612 .....GCAGAGAGAAAGAAAGAAAGAAAGAGAGAG 1641
1249 SerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnThrPr 1265
|||||:|||||
1642 GAGAAATGAAATGGCATTTCTGATAAATCTGAGGATGAGGCGCTCAGCC 1691
1265 oGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnA 1282
```



```
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154778.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0160741.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 176.00    Length: 1025
    Ratio: 0.424      Gaps: 44
    Percent Similarity: 40.488    Percent Identity: 19.220

alignment_block:
US-09-512-581-2 x AAC47245
Align seg 1/1 to: AAC47245 from: 1 to: 2638
395 GluArgThrLeuAspLysArgTrpArgValArgLys.....GluAlaMe 409
    ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
28 GAGCCAAACCGCAACAGACTACTTCTTGAAGCCCATCAGAGCCTAT 77
409 tMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaG 426
    | ||| :||| :||| :||| :||| :||| :||| :||| :|||
78 GGCTGGA.....AAGGAGAATGCTGGGG 100
426 lYLSAspAla.....AlaLysGlnIleAlaTrpIleLysAspLys 439
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 GPAAGAAACACAAAGACTGGCGAAGATGAGGATATGCTGAGCCAGAC 150
440 LeuLeuHisIleTyrTyrClnAsnSerIleAspArgLeuValG1 456
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 NATATGGAGATAGATGCTCAGATTAAAGAAAGATGATGAA...AAAGCTGA 197
456 uArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrGluA 473
    | :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 GACGGAAGATAAAGAGTCAGAGGTTAAGAAAAAATGAAGACAATGCTGAGA 247
473 tGMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaVal 489
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
248 CTCAAAAAATGGAAGAGAGAGGTGAGGTCACCAAGAGATGAGGACAAAGCA 297
490 LysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnVa 506
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
298 GAGGCTACCAACATGATGATCAAGATCCGATGCA..... 330
506 lLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValL 523
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 .....AAGAAAGACCAAACTGATGATGTGTGT. 357
523 ysAlaIlePheSerLysValMetValIleThrArgAsnLeuProaspPro 539
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358 .....TCAGTGGAAAGATACT 372
540 GlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAs 556
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
373 GTAATGAAGGAAACGTGGAATCTAAAGACAATAACTATGCCAAAGATGA 422
556 pGluLysIleArgLysGlnLeuGluVal..LeuValSerProThrCysSer 572
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
423 TGAATAAGAGACCAAGAGACAGACATATTACTG..... 454
573 CysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
455 .....AGGCAGACCACAAAAAGCTGGGA... 478
589 nProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuG 606
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
479 .....AGGAGGACATACAAC.....ATGAAGCTGACAAAGCAATGGAAC 518
606 luArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIle 622
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
519 AAAGATGGCA.....ATA 532
623 LysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluG1 639
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
533 CAGGAGACATCAAGAG.....GAAGGACACTGGTAGATGAAGACAAAGAG 578
639 yValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysValL 656
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
579 CACACATATGGATGAAAAAAGTGGAGAATGGGGATCAAAAAATAACAAGTGG 628
656 euSerPhe.....ThrHis 660
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
629 AGAATGTTGAAGAAAAAGAAAGAGATAAGAAAGAAAAATAAACAAAGAG 678
661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLe 677
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
679 GAAGTTGAGGGGCAAGAGCTGAGGTGATGAGTCA..... 714
```



```

6726 ...AGTTGCCAATCCCGAG.....:::|||||
1382 roSerProSerGlnPro 1387
6760 CCAGCCAGGCCAGCCA 6776

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA47245
seq_documentation_block:
ID_AA47245 standard; DNA; 2638 BP.
AA47245;
18-OCT-2000 (first entry)
Arabidopsis thaliana DNA fragment SEQ ID NO: 53101.
Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0123786.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
07-MAY-1999; 99US-0132487.
11-MAY-1999; 99US-0132863.
14-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144684.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
```


| | | | | | |
|------|---|------|---|------|---|
| 881 | aLaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluI | 898 | rGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer... | 1165 | S 1165 |
| 5093 | C.....TACAACA 5100 | 5093 | C.....TACAACA 5100 | 5093 | C.....TACAACA 5100 |
| 898 | leIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGlu 914 | 898 | leIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGlu 914 | 898 | leIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGlu 914 |
| 5101 | TCATCAACCTCTCAAGAGCAGCTGGCTGAGAGGCTCGGAAGGGAAC 5150 | 5101 | TCATCAACCTCTCAAGAGCAGCTGGCTGAGAGGCTCGGAAGGGAAC 5150 | 5101 | TCATCAACCTCTCAAGAGCAGCTGGCTGAGAGGCTCGGAAGGGAAC 5150 |
| 915 | CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSe 931 | 915 | CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSe 931 | 915 | CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSe 931 |
| 5151 | ACTAAGGAGATCCAGAG...TTCTCTGCGCCTGGCCAGGAGGTGGA 5197 | 5151 | ACTAAGGAGATCCAGAG...TTCTCTGCGCCTGGCCAGGAGGTGGA 5197 | 5151 | ACTAAGGAGATCCAGAG...TTCTCTGCGCCTGGCCAGGAGGTGGA 5197 |
| 931 | rArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaL 948 | 931 | rArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaL 948 | 931 | rArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaL 948 |
| 5198 | CAGAATGAACATGGGCTTG..... 5216 | 5198 | CAGAATGAACATGGGCTTG..... 5216 | 5198 | CAGAATGAACATGGGCTTG..... 5216 |
| 948 | ysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLys 964 | 948 | ysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLys 964 | 948 | ysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLys 964 |
| 5217 |CCTTCTCGGAGAGCATCAACCAAGCAAG.....GAG 5252 | 5217 |CCTTCTCGGAGAGCATCAACCAAGCAAG.....GAG 5252 | 5217 |CCTTCTCGGAGAGCATCAACCAAGCAAG.....GAG 5252 |
| 965 | AsnIleAsnVal.....ArgArgGluTyrLeuLysGlnHisAl 977 | 965 | AsnIleAsnVal.....ArgArgGluTyrLeuLysGlnHisAl 977 | 965 | AsnIleAsnVal.....ArgArgGluTyrLeuLysGlnHisAl 977 |
| 5253 | ATATGACCGCAGGCGCTGGCCCGAGGCCCGCCAGAGCTCTCAAGCTGGGAC 5302 | 5253 | ATATGACCGCAGGCGCTGGCCCGAGGCCCGCCAGAGCTCTCAAGCTGGGAC 5302 | 5253 | ATATGACCGCAGGCGCTGGCCCGAGGCCCGCCAGAGCTCTCAAGCTGGGAC 5302 |
| 977 | alaValSerGluLysLeuLeuSerLeu..... 986 | 977 | alaValSerGluLysLeuLeuSerLeu..... 986 | 977 | alaValSerGluLysLeuLeuSerLeu..... 986 |
| 5303 | AGCTCTCTCAGTAGAGCGGTACCAACTGAGAACAAAGTCCCGAGGCCCAAG 5352 | 5303 | AGCTCTCTCAGTAGAGCGGTACCAACTGAGAACAAAGTCCCGAGGCCCAAG 5352 | 5303 | AGCTCTCTCAGTAGAGCGGTACCAACTGAGAACAAAGTCCCGAGGCCCAAG 5352 |
| 987 |LeuProGluTyrValValPro...TyrThrIleHisLeu 998 | 987 |LeuProGluTyrValValPro...TyrThrIleHisLeu 998 | 987 |LeuProGluTyrValValPro...TyrThrIleHisLeu 998 |
| 5353 | ACTCTGGACATCAGCCAGAAATTTAGCTTACAGGGTCCCAACAAACACCTG 5402 | 5353 | ACTCTGGACATCAGCCAGAAATTTAGCTTACAGGGTCCCAACAAACACCTG 5402 | 5353 | ACTCTGGACATCAGCCAGAAATTTAGCTTACAGGGTCCCAACAAACACCTG 5402 |
| 999 |LeuAlaHisAspProAspTyrValLysValGlnAspIleG 1012 | 999 |LeuAlaHisAspProAspTyrValLysValGlnAspIleG 1012 | 999 |LeuAlaHisAspProAspTyrValLysValGlnAspIleG 1012 |
| 5403 | CGCTCCCGCTGCTCAGTGTAGACAA.....CGGTACCAAGATCTCCA 5446 | 5403 | CGCTCCCGCTGCTCAGTGTAGACAA.....CGGTACCAAGATCTCCA 5446 | 5403 | CGCTCCCGCTGCTCAGTGTAGACAA.....CGGTACCAAGATCTCCA 5446 |
| 1012 | uGlnLeuLysAspValLysGluCysLeuTrpPhe..... 1023 | 1012 | uGlnLeuLysAspValLysGluCysLeuTrpPhe..... 1023 | 1012 | uGlnLeuLysAspValLysGluCysLeuTrpPhe..... 1023 |
| 5447 | GGAGAAGCTGCTCATCTCAGAAAGCAGCTGTGTTCGCCAGGCCAAACCCAGC 5496 | 5447 | GGAGAAGCTGCTCATCTCAGAAAGCAGCTGTGTTCGCCAGGCCAAACCCAGC 5496 | 5447 | GGAGAAGCTGCTCATCTCAGAAAGCAGCTGTGTTCGCCAGGCCAAACCCAGC 5496 |
| 1024 |ValLeuGluIleLeuMetAlaLysAsnGluAsn 1034 | 1024 |ValLeuGluIleLeuMetAlaLysAsnGluAsn 1034 | 1024 |ValLeuGluIleLeuMetAlaLysAsnGluAsn 1034 |
| 5497 | TAGAGAAGTACAGAGCCATTTAAAGTAATCCCTGGTGAAGCAGACAGC 5546 | 5497 | TAGAGAAGTACAGAGCCATTTAAAGTAATCCCTGGTGAAGCAGACAGC 5546 | 5497 | TAGAGAAGTACAGAGCCATTTAAAGTAATCCCTGGTGAAGCAGACAGC 5546 |
| 1035 | AsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLy 1051 | 1035 | AsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLy 1051 | 1035 | AsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLy 1051 |
| 5547 | AGCAG.....ATCCAGTGGACCTTTCAGGACCTGGGCTATGAGAC 5587 | 5547 | AGCAG.....ATCCAGTGGACCTTTCAGGACCTGGGCTATGAGAC 5587 | 5547 | AGCAG.....ATCCAGTGGACCTTTCAGGACCTGGGCTATGAGAC 5587 |
| 1051 | sAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeuTyrThrV 1068 | 1051 | sAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeuTyrThrV 1068 | 1051 | sAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeuTyrThrV 1068 |
| 5588 | TTGTGGCCGAAGTGAGATGAAGCTGAAGCTGAGGAGAGACCACCCCTG 5637 | 5588 | TTGTGGCCGAAGTGAGATGAAGCTGAAGCTGAGGAGAGACCACCCCTG 5637 | 5588 | TTGTGGCCGAAGTGAGATGAAGCTGAAGCTGAGGAGAGACCACCCCTG 5637 |
| 1068 | aLysAspValAlaMetAsnIle.....IleMetSerLysSerThrThr 1082 | 1068 | aLysAspValAlaMetAsnIle.....IleMetSerLysSerThrThr 1082 | 1068 | aLysAspValAlaMetAsnIle.....IleMetSerLysSerThrThr 1082 |
| 5638 | AGTGTGAGGAGCAGGTACCTGAAGCCCTGGTGGTGAAGGCTTG 5687 | 5638 | AGTGTGAGGAGCAGGTACCTGAAGCCCTGGTGGTGAAGGCTTG 5687 | 5638 | AGTGTGAGGAGCAGGTACCTGAAGCCCTGGTGGTGAAGGCTTG 5687 |
| 1083 | TyrSerLeuGluSerProLysAspProValLeuProAlaAtrgPhePheTh 1099 | 1083 | TyrSerLeuGluSerProLysAspProValLeuProAlaAtrgPhePheTh 1099 | 1083 | TyrSerLeuGluSerProLysAspProValLeuProAlaAtrgPhePheTh 1099 |
| 5688 | TGCTGTGACCAAGGTTACCTGGACCCCTGCTCTG 5725 | 5688 | TGCTGTGACCAAGGTTACCTGGACCCCTGCTCTG 5725 | 5688 | TGCTGTGACCAAGGTTACCTGGACCCCTGCTCTG 5725 |
| 1099 | rGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluM 1116 | 1099 | rGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluM 1116 | 1099 | rGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluM 1116 |
| 5726 | CTCACCTGTGAAGACCTTGTGGAACAAGCCAG.....GAAG 5763 | | | | |


```
3792 .....TATCTAGG.....GGGAAGGAATT 3811
|||||
333 sValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysA 350
|||||
3812 AATGGCAGCATCTCAGGCATTCTCTTAACCAACAGCTGGCAGCACT 3861
|||||
350 sPheThrGluTyr.....LeuLysVal 357
|||||
3862 CTGTAGGCCCCCACCATTGAGAGCAAACTGACCAAGTTCTACGAGATG 3911
|||||
358 ArgSerHisAspProGluGluAlaLys.....ArgHisAspValIleVa 372
|||||
3912 CCCTCTCGAGACGACAGCACCCTGCCTGACTGCCAGAGAGGAGCCAGCAT 3961
|||||
372 lSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisL 389
|||||
3962 ACCCCGGTCTACATTAGGAGACTCAGACACAGTTGCGAGGCTGGAGAAAG 4011
|||||
389 euleuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArg 405
|||||
4012 AACTGAGCAATGCCAGGAGGAG....CTGAGCTCATGCCCAAAAGAA 4058
|||||
406 LysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuG1 422
|||||
4059 AGAGAAAGCCAGATAGATTGTCTGCCCTGCAGTCCATGATGCGCTGCA 4108
|||||
422 n.....SerAlaAlaGlyLysAspAlaLysGlnI 433
|||||
4109 AGAGGAAGAGCTGCAGGTCGAGGCTGCTGACTTGGAGTCCCTGACCAGGA 4158
|||||
433 leAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
|||||
4159 ACATACAGATAAAGAGACCTCATAAAGACCTGCAAAATGCAACTGGTT 4208
|||||
450 Asp.....AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVa 464
|||||
4209 GACCCCTGAAGATATGCCAGCATGAGGCGCTGACCCAGAGGCTTACT 4258
|||||
464 l.....ProHisAsnLeuGluThrG 472
|||||
4259 TCTTCGGGAAAAAGTTGCTTCACTGGAAACCCAGGTCAGGAAGGCTCAG 4308
|||||
472 lu.....ArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
|||||
4309 AGACAGGAGACAACTGCTCTGCTGATGTAGAGGACTAGTGGATGAA 4358
|||||
487 AsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuAr 503
|||||
4359 CGGAGT.....CGGCTCAACGAGGCCCTGCAAGCTGAG.....CG 4393
|||||
503 gHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspA 520
|||||
4394 GCAGCTCTACAGAGCGCTGGTCAAGTTCCATGCCCAACAGAGATCTCTG 4443
|||||
520 laSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
|||||
4444 AGAGACCCGAACTCTG..... 4460
|||||
537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnVal.. 552
|||||
4461 .....CAGGTGGA 4468
|||||
553 LeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerP 569
|||||
4469 ACTGGAAGGGCCGAGGTGTAGCGACTGCAGTAGAA..... 4505
|||||
569 roThrCysSerCysLysGlnAlaGluCysValArgGluIleThrLys 585
|||||
4505 ..... 4505
|||||
586 LysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLy 602
|||||
```


antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;
 antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;
 atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
 eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
 reperfusion injury; atopic dermatitis; diabetes insipidus;
 conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
 arterial restenosis; ankylosing spondylitis; transplant rejection;
 graft versus host disease; ss.

Rattus sp.

Key Location/Qualifiers
 CDS 1746..8726
 /*tag= a
 /product= "Myomegalin"
 /note= "this region is referred to as SEQ ID NO: 3
 in the specification"

W0200027861-Al.

18-MAY-2000.

12-NOV-1999; 99WO-US26860.

12-NOV-1998; 98US-0108255.

(STRD) UNIV LELAND STANFORD JUNIOR.

Conti M, Pahlke G;

WPI; 2000-376479/32.

P-PSDB; AAY71158.

Polynucleotide encoding a phosphodiesterase (PDE) interacting
 polypeptide, useful for diagnosis and treatment of asthma, cystic
 fibrosis, Crohn's disease, and rheumatoid arthritis -

Disclosure; Fig 2; 77pp: English.

The present sequence is a cDNA encoding a phosphodiesterase (PDE)
 interacting protein, myomegalin from rat skeletal muscle cDNA library.
 The protein modulates the functions and properties of PDEs, specifically
 cAMP-PDEs, and also targets them to specific subcellular compartments.
 Rat myomegalin has at least four differently sized transcripts, two of
 which are expressed in heart (7.5 and 5.9 kb), two in skeletal muscle
 (7.5 and 4.3 kb) and one in testis (2.5 kb). The present sequence
 can be used in the diagnosis and treatment of disease conditions
 associated with PDE activity. The diseases include asthma, cystic
 fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
 granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
 septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
 inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
 distress syndrome, diabetes insipidus, allergic rhinitis, allergic
 conjunctivitis, vernal conjunctivitis, arterial restenosis,
 atherosclerosis, inflammatory diseases associated with irritation and
 pain, rheumatoid arthritis, ankylosing spondylitis, transplant
 rejection and graft versus host disease, disease conditions associated
 with hypersecretion of gastric acid, and disease conditions in which
 cytokines are mediators.

Sequence 9679 BP; 2549 A; 2626 C; 2636 G; 1867 T; 1 other;

alignment_scores:
 Quality: 177.00 Length: 1539
 Ratio: 0.270 Gaps: 72
 Percent Similarity: 42.625 Percent Identity: 18.583

alignment_block:
 us-09-512-581-2 x AAD00768 ..

Align seg 1/1 to: AAD00768 from: 1 to: 9679

22 lIeSerAspLysIleSer.....LysGluGluMetValArgAr 34
 |||:|||||:|||||
 3219 ATAGATGAAAGTTCTCTCTCTAGCAAGAAAGGACAAAGAACTCGCGCA 3268
 34 gLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGluG 51
 |||:|||||:|||||
 3269 GCTCGGGTGTCTGTGAGG.....GACCGAGACCATGACTTAGAGA 3309
 51 luGluLys...GluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
 |||:|||||:|||||
 3310 GACTCGGTGTCTCTCTGCTGCCAATGAAGTACCATGCAAGATGAGAG 3359
 67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCy 83
 |||:|||||:|||||
 3360 AGTCCTCTGAGGCCAGAGGCTGGAAGTGGAGCAGATTAAATGCCACCTG 3409
 83 sIeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerP 100
 |||:|||||:|||||
 3410 C..... 3410
 100 roAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116
 3410 3410
 117 LeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGl 133
 3411CA 3412
 133 uAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAsps 150
 |||:|||||:|||||
 3413 AAACCTCCAGTGGTTGAAGGAA.....GAATTGGAA... 3443
 150 erAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn 166
 |||:|||||:|||||
 3444ACCAAGTTT..... 3452
 167 AsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSe 183
 |||:|||||:|||||
 3453 ...GCCCATGGCAGAAGAACAGAGGAGAGCATCATTCAGCAGTACAGAC 3499
 183 rIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 200
 |||:|||||:|||||
 3500 ATCTCTGATGACAGAACAAAGAGTAGAGGATCTCAGTGCACACTTGC 3549
 200 euValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAsp 216
 |||:|||||:|||||
 3550 TCCACAAACTTGGACCGCCGAGAGTGAAGTA.....GCTGAGGAG 3590
 217 LeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIl 233
 |||:|||||:|||||
 3591 CTGTGCCAGCGCTGCAGCGGAAG..... 3614
 233 eThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSeRa 250
 |||:|||||:|||||
 3615GAAAGGGTCTG.....CAGGACCTTCTGAGTG 3642
 250 sPLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAsp 266
 |||:|||||:|||||
 3643 ATCGGAACAACAAGCCATGGAGCAGACGATGGAGTCCAGGGA..... 3686
 267 SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSe 283
 |||:|||||:|||||
 3687CTGCTCCAGTCGATG.....GG 3703
 283 rAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetP 300
 |||:|||||:|||||
 3704 CACCGGGGAACAGAAAGACAGGCTGTTCAGAAAAAATGGTACAGCCT 3753
 300 heGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316
 |||:|||||:|||||
 3754 TCATGGAAGAAACTCGGAATTACAGGCCCTCGCGCAG..... 3791
 317 CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCy 333


```

3029 AAAAGTAATGAGTGTGCAGAGATGCTAA...CAAAAAGTAGAGGAGCT 3075
      ::::: |||::: ||| ::::: ::
832 tValArgTrrpLeuLeuGlyMetLysAsnHisSerLys..... 845
      ::::: |||::: ||| ::::: |||
3076 TACTAGG.....CAGAATAATGAACCAAAATTAATGATGC 3110
      ::::: |||::: ||| ::::: |||
846 .....SerGlyThrSer 849
      ::::: |||::: ||| ::::: |||
3111 AGAATTCAGGCAGATCAGCAATATAGTTTACAGAAACAATATC 3160
850 ThrLeuArgLeuLeuThrThrLeuHisSerAspLys..... 862
      ::::: |||::: ||| ::::: |||
3161 CAGCAAGATGTGAAGCAGCTGAAATTCAGGCAGATGGTTTAAAGATCA 3210
863 .....AspLeuThrGluGlnGlyLysIleSerLysP 873
      ::::: |||::: ||| ::::: |||
3211 GCTACGAGCAGCAAAATGAACACTTGCATAAACAGACAAAACAGCAGG 3260
873 roAspMetSerArgLeuArg.....LeuAlaAlaGlySer 884
      ::::: |||::: ||| ::::: |||
3261 ATTTTCAACAAAATTAATGCTAGAGAACACCTGGCGCAAAAGTCAA 3310
885 AlarLeuValLysLeuAlaGlnGluProCysTyrHisGlnIleLeuThrLe 901
      ::::: |||::: ||| ::::: |||
3311 AATTGGTGAAGTGAATTAAGCAAAAGTGTGACCAACAGCAACATTTATCAT 3360
901 uGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys..... 915
      ::::: |||::: ||| ::::: |||
3361 CCAGATACCAAGAAAGATTAAGAAATCTGAATCGGAACCTGAATGCTT 3410
916 .....TyrGlnValArgGlnVal 921
3411 CCAAGAGAGACCGCGCGGGGAGCAGAAAGTTACGCTACACAAAGCT 3460
922 PheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuG1 938
      ::::: |||::: ||| ::::: |||
3461 CAGGTGCAAGAGTTAAATAACAGGTTTCAAAAAAGTACACAGCAAGTTTACA 3510
938 uTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgA 955
      ::::: |||::: ||| ::::: |||
3511 CTTAAAGACCATTA.....GAGGAGCAGA 3533
955 rGAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
      ::::: |||::: ||| ::::: |||
3534 TGACCCACAAGAGATGTTCTGTTCCAGGAAGAAATCGGTAA... 3577
972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuPr 988
      ::::: |||::: ||| ::::: |||
3578 ...TTCAACAATCAGCAGAGAGGAGTTTCGGAAGAAGATGGAATAATTAAT 3624
988 oGluTyrValValProTyrThrLeuHisLeuLeuAlaHisAspProAspT 1005
      ::::: |||::: ||| ::::: |||
3625 GGAGTCCAAAGTCTATCACTGAAATGATATTTCCAGGCATTAGGCTTGACT 3674
1005 yrValLysValGln..... 1009
      ::::: |||::: ||| ::::: |||
3675 TTGTGCTCTTCACAGAAACTCTAGAGCCCAAGAAATGCTAAGCTT 3724
1010 .....AspIleGlu...GlnLeuLysAspValLysG1 1019
      ::::: |||::: ||| ::::: |||
3725 TGTGAACAACAACAPTAAAGAACTTGAAGACAGCTTCAACAGTATCGTGA 3774
1019 u.....CysL 1021
      ::::: |||::: ||| ::::: |||
3775 ACAATGCAGCAGGAGCAGCAGATGGAAGCAAAATCATTAACCAAAATGTC 3824
1021 eurrPpPheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHis 1037
      ::::: |||::: ||| ::::: |||
3825 AGAAACTTGAGGATGAGCTGATGCCAGAGCGGTGAG..... 3862
1038 AlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaG1 1054
      ::::: |||::: ||| ::::: |||

```

```

3863 .....GTTGAAAACCTGAAGCAAAAAATGGACCAACA 3894
1054 nGlyProAspAlaLysMetAsnGluLysLeuTyrThrVal..... 1068
      ::::: |||::: ||| ::::: |||
3895 G.....ATCAAGAGCATGAACATCAATTAGTTTGTCTCCAGT 3932
1069 .....Cys.....Asp 1070
      ::::: |||::: ||| ::::: |||
3933 GTGAATTCAAAAAAGACAGCAGCAAGAGCTGTACCTTCAAAACCATAT 3982
1071 ValAlaMetAsnIle.....IleMetSerLys 1079
      ::::: |||::: ||| ::::: |||
3983 TTTGAGATCAGACTGAAGAGAGTCCAGCACTCTGGAGAGCTGTCTCTAG 4032
1079 sSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaA 1096
      ::::: |||::: ||| ::::: |||
4033 AAACACTGGACACTTCAACCAACACCATCCCTCTCTGTG.....A 4076
1096 rgPhePheThrGlnPro.....AspLysAsnPheSerAsnThr 1108
      ::::: |||::: ||| ::::: |||
4077 GATGGACTCAAGAACCAACAGCCATTGGAAGAGAAGTGGCAGCATCGGTT 4126
1109 LysAsnTyrLeuProGluMetLysSerPhePheThrProGlyLysPr 1125
      ::::: |||::: ||| ::::: |||
4127 GTTGAACAGATACCAAAAGAGTCCAA...TTCAGCCACACAGG..... 4168
1125 oLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaG 1142
      ::::: |||::: ||| ::::: |||
4169 .....GTCCTCACTCGAAGAAAGAGA 4187
1142 lLysGln.....SerGlnThrLysSerSerArg 1151
      ::::: |||::: ||| ::::: |||
4188 AAGCCAGCAGCTGTACTCTGAGTACTTTTCTCAGACAAGCAGGATTA 4237
1152 MetGluThrValSerAsnAlaSerSerSerSerSerSerProG1 1168
      ::::: |||::: ||| ::::: |||
4238 CAGATAACTTTT.....GATGAGACAAACCCATTACA..... 4270
1168 yArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnG 1185
      ::::: |||::: ||| ::::: |||
4271 .....AGACTG.....TCTGAATTTGAGAAGATAAGAGACC 4301
1185 luAspTyrThrMetSerSerProLeuProGlyLysLysSerSerLysArg 1201
      ::::: |||::: ||| ::::: |||
4302 AAGCCCTGAACAATCTAGACCA...CCTGTAGGTATCAAGATAAGCCA 4348
1202 AspAspSerAspLeuValArg.....SerGluLeuGluLysProArgG1 1216
      ::::: |||::: ||| ::::: |||
4349 TGTGAATGGAAGTGTGAAGTTTTCACACCTTAGAGATAGTAAGAA 4398
1216 yArgLysLysThrProValThrGluGlnGluLysLeuGlyMetAspA 1233
      ::::: |||::: ||| ::::: |||
4399 CAAGCAGATGATGATGCATACAGAA..... 4423
1233 spLeuThrLysLeuValGlnGluGlnLysPro 1243
      ::::: |||::: ||| ::::: |||
4424 ..GTCACACATTAACAAAGAAAGAACCCA 4453

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAD00768

seq_documentation_block:

ID AAD00768 standard; cDNA; 9679 BP.

XX AAD00768;

AC AAD00768;

DT 08-SEP-2000 (first entry)

DE Rat phosphodiesterase interacting protein, myomegalin cDNA.

KW Rat; phosphodiesterase interacting protein; myomegalin; PDE; CAMP-PDE;
 KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
 KW antiinflammatory; antipruritic; dermatological; shock;
 KW analgesic; immunosuppressive; antiulcer; vasotropic; antiarthritic;

| | | |
|------|---|------|
| 252 | SerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHis | 268 |
| 1706 | CTTCAGCGCTCAGAAGCAACAGTCTTGAGAAATAGCAAACTTACAGAAA | 1755 |
| 268 | sLeuLeuLeuSerValLeuProGlnLeuGluPheLeuLeuLeuLeuSerAsnA | 285 |
| 1756 | G.....ATAGTGTGGTGGAAAGAAAGTGTAGCTGAACATAAG..... | 1792 |
| 285 | sPAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGly | 301 |
| 1793 |AAACAAAAGTCCCGACTAGTAGGAGAAACATTCGAAAGTCCAGGAG | 1837 |
| 302 | AlaLysAspSerGluLeuAlaSerGlnAsnIleProLeuTrpGlnCysTyr | 318 |
| 1838 | GCTGCAGAAAATGAATTGAGAAACACGACAGAGA..... | 1870 |
| 318 | rLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValL | 335 |
| 1870 | | 1870 |
| 335 | ysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeu | 351 |
| 1871 |AATGTAGAAGATATCTCTCTGCAG...1894 | 1894 |
| 352 | ThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHis | 368 |
| 1895 |AAGATAAGGCGCTGAAGTGAAGCCCAAGCAGTACCCGAC | 1932 |
| 368 | sAspVal.....IleValSerIleValThrAlaAlaLysLysAspIleL | 383 |
| 1933 | GGAACTTGAAACCATGTGACAGACAGGAGGACCGCTGAAGAGAGA... | 1978 |
| 383 | eLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAsp | 399 |
| 1979 |CTGGAGCGGGTGACGACGTCACCATAGAG | 2008 |
| 400 | LysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyr | 416 |
| 2008 | | 2008 |
| 416 | rLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnI | 433 |
| 2009 |GCCGAGCTAAAGAGCTGCC.....2029 | 2029 |
| 433 | IeAlaTrpIleLysAspLysLeuHisIleTyrTyrGlnAsnSerIle | 449 |
| 2030 |GTGGAAGAGAACCCTCTGAAT.....TTTCGAATTCAGTTG | 2065 |
| 450 | AspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis | 466 |
| 2066 | GAGGAAAACACCTTTACCAGACGACAACTGGGAAGATCATCTTAAAGAAA | 2115 |
| 466 | sAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaT | 483 |
| 2116 | AGATTTA.....2122 | 2122 |
| 483 | hrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGln | 499 |
| 2123 |AGTCTCAATGATTTGGACGACAAACAAA | 2149 |
| 500 | AsnLeuLeuArgHisGlnValLys.....AspLeu | 509 |
| 2150 | ATAAATAATGGAAGAAATTAAGAAGAAAGAGACAAATGAGGAAGAACT | 2199 |
| 509 | uLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIleP | 526 |
| 2200 | CTTGAAGCTGATAAGACGACAGTCGAAAAGACCTTGA.....T | 2237 |
| 526 | heSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAla | 542 |
| 2238 | TTTCAGAAA...CAGGTACGACAGAAACAGTTTGAAGAAAAGACGACGAAAT | 2284 |

543 Gln...AspPheMetLysLysPheThrGlnValLeuLeuGluAspSerGluLys 558
::: :::: ::::: |||::: |
2285 GAATTTGGAGCAAGAAGAAAATAACTCAA..... 2314

558 stleArgLysGlnLeuGluValLeuLeuValSerProThrCysSerCysLysG 575
|||::: :::: ::::: |||::: ||| |
2315 .ATTCCAGTATACGTAGAGAAATGCATTGCCCGCTGTGCCGATCACAC 2363

575 InAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLys 591
|||| :::: ||| :::: ||| |
2364 AGGCTCATCATGC..AGGCAGCTAACG.....GCTCTCCAGCAA 2401

592 GlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgII 608
::: :::: ||| :::: ||| |
2402 GAACATGACAAGCAGAAAGCAGAGAAACTCAACAGCAGGTAGTAGTA 2451

608 eaLaProValHisIleAspThrGlu...SerIleSerAlaLeuIleLysG 624
::: :::: ||| :::: ||| |
2452 AACAGCTGCCAATAGAAAGCGCTGAACAGACATCAGAGAGCTGACATATG 2501

624 InValAsnLys...SerIleAspGlyThrAlaAspAspGlyAspGluGly 639
::: ||| :::: ||| :::: ||| |
2502 ACTTAAATGCCCTCCAGCTTGAAAAACGTCATCTGAGGAAAG..... 2545

640 ValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLe 656
|||| :::: ||| :::: ||| |
2546GCTCGTTTGTCTAAAA..... 2560

656 uSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerL 673
::: :::: ||| :::: ||| |
2561GATAAAGCTAGATGAACCAAAATAATA 2585

673 euLeuAlaCysLeuLysMetAsp..... 680
||| ||| ||| ||| ||| |
2586 CACTCAGATGCTTAAGTTGGAGCTGGAAGGAGGATCAGGCGGAGAAA 2635

681AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnTh 695
||| ||| ||| ||| ||| |
2636 GGTTATTCTCAACAACCTCAGAGAGCTTGGTAGGCAATTGAATCAAAACC 2685

695 rClySerLysIleGluGlu.....AspPheProHisI 706
||| ||| ||| ||| ||| |
2686 AGGT...AAAGCTGAAGAGAGCCATCGAAGAGCTAGTGTCTCAAGAAA 2732

706 leArg.....SerAlaLeuLeuProValLeuHisLysSerLys 719
||| :::: ||| ||| ||| ||| |
2733 TAAAGCGCAATTATCAGTTAGCAATTAGCAATCTTAATCAT.....GAA 2776

720 LysGlyProProArgGlnAlaLysTyralaIleHisCystIleHisAlaII 736
|||| :::: ||| ||| ||| ||| |
2777 AAAGGGAACCTACAAAGAGAGTAGTACAGAGAATCAACAGGCACATGCTGT 2826

736 ePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisL 753
: ||| :::: ||| :::: ||| |
2827 A.....GCTGAGAGAATATTTCAGCATTTTAAATTCACAAATTCATT 2867

753 ysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuValThr 769
: ||| :::: ||| ||| ||| ||| |
2868 CTITTCGAGATGAGAGAAGAAATTAGAAGAGCTACAAATCTGCCAGAGAAA 2917

770 lleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaIaIaProTrpLy 786
|||| :::: ||| ||| ||| ||| |
2918 TCAGATCATCTA.....AAGAACAATTTTCAG.....AA 2946

786 sSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgL 803
||| :::: ||| ||| ||| ||| |
2947 AAGCCAT.....GAGCAGTGTCTTCAAAATATCAAAAG 2978

803 euProGlyLysLysThrThrLysLysLeuTrpValProAspGluGluValSer 819
||| :::: ||| ||| ||| ||| |
2979 CTGAAAAAGAAAAATAGTAATAATCCAAGGCTCAATGAAGAAATTCGAG 3028

820 ProGluThr.....MetValLysIleGlnAlaIleLysMetMe 832


```

1873 ATCACTGACCTG.....CAGAGCTCCATCTC 1898
249 rAspLeuSerGlu.....Hisv 255
|:|||||:
1899 CAACTCAGCCAGGCCAAGAGAGCTGGAGCAGGCTCCAGGCTCATG 1948
255 alPheAspLeuLeuGluLeuTyraSniLeAspSerHisLeu..... 269
|:|||||:
1949 GGGCCCGGTTGACTGCCAGGTGGCTCTCTGACCTCTGAGCTCACCACA 1998
270 LeuLeuSerValLeuProGlnLeuGluPheLys.....LeuLysSe 283
|:|||||:
1999 CTAATGCCACCATCCAGCAACAGCACTAAGAACTGGCTGGCTGAAGCA 2048
283 rAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetP 300
|:|||||:
2049 GCAGGCCAAGAGAACAGGCCCTGCTAGCACAGACCCTC..... 2088
300 heGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316
|:|||||:
2089 .....CAACAGCAAGAACAGGCTCCAGGGC..... 2115
317 CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCy 333
|:|||||:
2116 .....CTCCGCCACCAGGT 2129
333 sValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysA 350
|:|||||:
2130 GGAGCAGCTAAGCAGTAGCTGAAGCAGAGGAGCAGCAGTGT...AAGG 2176
350 spLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaLle 366
|:|||||:
2177 AGGTACGGAG.....AAGCAGGAGGCAACT 2202
367 ArgHisAspValIleValSerIleValThr..... 376
|:|||||:
2203 AGCAGGAGCATGCCAGCAACTGGCCACTGCTGCAGAGGAGCAGAGGC 2252
376 ..... 376
2253 CTCCTTAAGGAGCGGGATGGGCTCTAAGCAGCTGGAGGCACTGGAGA 2302
377 .....AlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeu 390
|:|||||:
2303 AGGAGAAGGCTGCCAGCTGGATTTCTGCAGCAGCAACTTCAGGTGGCT 2352
391 AsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGl 407
|:|||||:
2353 AATGAAGCCGGGACAGTGCACAGCTCAGTGACACAGGCCCGCCAGCGGA 2402
407 uAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr...AlaLeuGlnS 423
|:|||||:
2403 GAAG.....GCAGAGCTGAGCGGAAGGTGGAGGAACCTCCAGG 2440
423 erAlaAla.....GlyLysAspAlaAlaLysGlnIle 433
|:|||||:
2441 CTTGTCTTGAGACAGCCGCCAGACACATGATGAGCCCGCCAGGTT 2490
434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
|:|||||:
2491 GCAGAGCTAGTTGACGTGCGGTCTGAGCAGCAAAAAAGCAACTGAGAA 2540
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467
|:|||||:
2541 AGAAAGGTTGGCCAGAGAGGACCATGCTCCAGGAGCAGCTCCAGGCC 2590
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
|:|||||:
2591 TCAAGAGCTCTTGAAGTCAACAAG.....GCAGC 2622
484 LeuAsp.....LeuAsnAlaValLysAlaLeuAsnGluMetTrpLys 497
|:|||||:
2623 CTTGAAGAGGAGAAAGCGGCTGCAGATGCCCTGGAAGAGCAGCAGCG 2672

```

```

497 sCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLeu 514
|:|||||:
2673 TTGTATCTCTGAGCTGAAGCAGAGACCCGACCTGGTGGACGCATATA 2722
514 ysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
|:|||||:
2723 AGCGGGAACAAAGGAG..... 2739
531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
2739 ..... 2739
547 slsYPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuG 564
|:|||||:
2740 .....CTGGAAGAAGAGAGGGCTGGCGCAAGGGCTGG 2773
564 luValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
|:|||||:
2774 AGGCTCGATTACAGCAGCTTGGGAGGCCCATCAGGCTGAG..... 2814
581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPh 597
|:|||||:
2815 ACTGAAGTCTCGCG..... 2829
597 eLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleA 614
|:|||||:
2830 .....CGGAGCTGCCAGAGCCCTGGCTGCCACGACACAG 2866
614 spThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
|:|||||:
2867 CTGAGAGTGAGTGTGAGCAGCTCGTCAAGAACTAGCTGCTGCGCTGAG 2916
631 GlyThrAlaAspAspGluGlyValProThrAspGlnAlaIleAr 647
|:|||||:
2917 CGGTATGAGGATAGCCAGCAAGAGGAGGACAGTATGGCCCATGTCCA 2966
647 gAlaGlyLeuLeuLeuLysValLeuSerPheThrHisProIleSerP 664
|:|||||:
2967 GGACAGCTGATGACTTTGAG..... 2988
664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
|:|||||:
2989 .....GAGGAATGTGAGAAGCCCGCCAGGAGCTGCAGGAGGCA 3027
681 AspGluLysValAla.....GluAlaAlaLeuGlnIlePheLy 693
|:|||||:
3028 AAGGAGAAGGTGCGAGGCATAGAATCCACAGCAGCTCCAGATA.... 3072
693 sAsnThrGlySerLysIleGluAspPheProHisIleArgSerAlaL 710
|:|||||:
3073 .....AGCCGCGCAGCAAGCAACTAGCTGAGCTCCATGCCCAACC 3112
710 euLeuProValLeuHisLysLysLysGlyProProArgGlnAla 726
|:|||||:
3113 TGCCAGAGCACTCCAGCAGGTCCAGAGAAG...GAAGTCAGGCGCCAG 3159
727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGl 743
|:|||||:
3160 AAGCTTGCAGATGAC.....CTCTCCACTCTGCAGGAGAAA 3194
743 nPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnL 760
|:|||||:
3195 GATGGCTGCC.....ACCAGCAAGAGGTGGCCGCT 3226
760 euGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776
|:|||||:
3227 TGGAGACCTTGGTG..... 3240
777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIl 793
|:|||||:
3241 .....CGCAAGGCGAGGTGAGCAGCAGGAAACAGCTCCCGGGAGTT 3281

```



```

|||||: 3042 AGTCAAGGAGCCTCGGAGGCGAGGAGACAGACAGCCGAG. 3081
809 hrLysLeuTrpValProAspGluGluValSerProGluThrMetValLys 825
|||||: 3082 .....TGGCTG.....GAAGAGCAACAGGAGCGCCCTTCGCAGC 3117
826 IleGlnAlaIleLysMetMetValArgTrpLeuLeuGluMetLysAsnAs 842
|||||: 3118 ACACAGCAGCG.....CTCAGGCTATGGAGCGGA 3149
842 nHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuH 859
: 3150 GGCAGACAGATGGCAATAGCTGGAACGGCTCGCGCGCGCTGATGG 3199
859 isSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
: 3200 AGAGCCAGGCGCAGCAGCAGGAGCGGTGGGAGCAGGAGGAGGGGTG 3249
876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnI 892
: 3250 GCGCGGCTGACCCAGAGCGGCGCGTGCAGGCTGACCTTGCCCTGGA 3299
892 uProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaL 909
3300 G..... 3300
909 euAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLys 925
3300 ..... 3300
926 LeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaI 942
3300 ..... 3300
942 eCysAlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaA 959
3301 .....AAGCGGCCAGAGCAGAGCTTG 3322
959 rgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
3323 AGATGCGGCTGCAGAGAGCGCCTCAACAGCAGCGT.....GTGGAG 3363
976 HisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValVa 992
3364 TTGCTACCTCTGAAGAGGCACCTGCTATGCTGACCGCTGACGGAAGAAGAA.. 3411
992 lProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValG 1009
3412 .....GGCAGGACCCAGGAGTGGCCCAAGCTTC 3439
1009 InAspIleGlu.....GlnLeuLysAspValLysGluCysLeuTrpPhe 1023
: 3440 GTGCTCTGGAGGCGCCAGACATAAAGAGCTGGAGGAACCTTCGCACAAAC 3489
1024 Val.....LeuGluIleLeuMetAlaLysAsnGluAsnSerHisAl 1038
||| 3490 GTAAGCAACTGAAGGAACAGCTGGCTAAGAAAGAAAGAGGACCGCATC 3539
1038 aPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnG 1055
: 3540 TGGCTCAGGAGCCCAACTCTGAGGCTGCTGCGCAGGACAGAG. 3579
1055 lyProAspAlaLysMetAsnGluLysLeuTyrThrValCysAspVal 1071
||| 3580 ..CCAACAGCCCAAGCTGAGGCACTGCGGCGCAGAGGTGAGCAAGCTG 3627
1072 AlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerPr 1088
: 3628 GACAGCAATGCCAGAGCAGGAGCAGGCTGACAGCCTGGAAACGCAG 3677
1088 oLysAspProValLeuProAlaArgPheThrGlnProAspLysAsnP 1105
: 3678 CCTCGAGGCTGAGCGGCCCTCCCGG.....GCTAGCGGGGACAGTGTCTC 3721
1105 heSerAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThr 1121
: 3722 TGGAGACTTCGACGGCCAGTTCAGAGGAGAGGCCAGGAG..... 3762
1122 ProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
|||||: 3763 .....CTAGGCGACAGTCAGAGTGCCTT 3785
1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThr 1154
: 3786 AGCCTCGGCCCAACGGGAGTGTGCTGCTTCCGACCAAGGTACAAGACC 3835
1155 ValSerAsnAlaSerSerSerSerSerSerSerProGly..... 1168
: 3836 ACAGCAAGGCTGAGATGAGTGAAGCCAGGTCGCCCGGGCCGCCAA 3885
1169 ArgIleLysGlyArgLeuAspSerSerGluMetAsp..... 1180
: 3886 GAGCTGAGAGGAAATACCTCATCAGCAGCTTGGAGGAGGAGTGTCTC 3935
1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysL 1197
||| 3936 CATCTGAA.....TCGCCAGGTCTCGAGAAGG 3964
1197 ySerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLys 1213
: 3965 AGGGGAGAGCAAGGAGTTGAAGCGGTGATGCGCCGAGTCAGAGAAG 4014
1214 ProArgLysLysLysThrProValThrGluGlnGluLysLeuG 1230
: 4015 AGCAG.....AAGCTGGAGGAGAGGCTG... 4038
1230 yMetAspAspLeuThrLysLeuValGlnGlnLysProLysGlySerG 1247
: 4039 .....CGCCTGTGTCAGGAGAGACAGACCCAGCAACAGTG 4072
1247 InArgSerArgLysArgGly.....HisThrAla 1256
: 4073 CCAGAGCTGCAGAACCGCAGCTCTCTCGCGGAGGAGGTGCAGAGCTC 4122
1257 SerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAs 1273
: 4123 CGGAGGAGGCTGAGAAACAGCGGTGCTTCAGAGAACCTGCGGAGGA 4172
1273 pIle.....LeuGluAsnGluAspGluGlnAsnSerProLys. 1286
: 4173 GCTGACCTCACAGGCTGAGCGTGGGAGGAGCTGGGCCAAGAATTGAAGG 4222
1287 LysGlyLysArg.....GlyArgProProLysProLeuG 1298
: 4223 CGTGGCAGGAGAAAGTTCTCCAGAAAGAGCAGGCGCTCTCCACCTGCAG 4272
1298 yGlyGlyThrProLysGluPro.....ThrMetLysThrS 1311
: 4273 CTCGAGCACACCCAGCACAGGCCCTGGTGAGTGAGCTGTGCCAGCTAA 4322
1311 erLysLysGlySerLysLysSerGlyProProAlaProGluGluGlu 1327
: 4323 GCACCTCTGCCAGCAGCTGCAGGCGCAGGCGCTGCCGAGA..... 4366
1328 GluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLy 1344
: 4367 .....AACGCCACCTGAGGAGCTGGAGCAG 4392
1344 sGlnHisArg..... 1347
: 4393 AGCAAGCAGCCCGCTGGGGGAGCTGCGGGCAGAGCTGTGCGGGCCAGCG 4442
1348 .....ValSerArgAlaGlnGlnArgAla 1356
: 4443 GGAGCTTGGGAGCTGATTCTCTCTGCGGAGAGGTGCGAGCAGGAGC 4492

```


[illegible]

```

3553 ATTAAATGAGGCA.....:|||||:
409 eWetGlyLeuAlaGlnIleTyrLysTyrAlaLeuGlnSerAlaAla 425
3585 TTTTAAATTGGAGATGATTAAA.....:|||||:
426 GlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHi 442
3610 .....:|||||:
442 sIleTyrTyrGlnAsnSerIleAspAspArgLeuValGluArgIleP 459
3611 .....:|||||:
459 heAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
3645 AGATGAA.....CCGGTCAAAAGAG.....:|||||:
476 CysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLe 492
3668 .....GTCAAAAAGAACTGTTAGTATTAT 3693
492 uAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAsp 509
3694 TGAAGAAATG.....GAAGAAATATTCTAGATG 3722
509 euleuAspIleuIleLysGlnProLysThrAspAlaSerValLysAlaIle 525
3723 TATTAGAGGAAGAAAAGAAATTTAAACAGACAGATGATAGATGCAGTA 3772
526 PheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
3773 GAAGAAATCATTAGAAATATCTTCAGATTCTAAAGAGAACTGAATCTAT 3822
542 acIlnAspPheMetLysLysPheThrGlnValLeuGluAsp.....:|||||:
3823 TAAAGATAAAGAAAAGATGTTTCTACTAGTTGTTGAAGAAGTTCACAGACA 3872
556 .....AspGluLysIleArgLysGlnLeuGluValLeuValSerPro 569
3873 ATGATATGATCAAGATGTTGTGAGAAGATTTTGAAGTGAAGAAATATGCGA 3922
570 ThrCysSerCysLysGlnAlaGluCysValArgGluIleThrLysLys 586
3923 GAGGAGTTAATCAAGAGTGCTGTGAA...ATGAATGCATTACTACGAA 3969
586 sLeuGlyAsnProLysGlnProThrAsnProPhe...LeuGluMetIleL 602
3970 ACTTATTGAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAA 4019
602 yspPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIle 618
4020 AA.....GATGTGGAAGAAATTA 4036
619 SerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAs 635
4037 AAGAANTTAAAAAGACATTATCAGAGATTTCAAGAAATAATAGATGC 4086
635 pGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluL 652
4087 AAAAGATGAT.....ACAT 4100
652 euleuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGlu 668
4101 TAGAAAAAGTTATTGAAGAGGAACATGATATAAGC.....ACG 4138
669 ThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAl 685
4139 ACGTTGGATGAAGTTGTAGAATTAAAGATGTCGAAGAAGCAAGATCGCA 4188
685 agIluAlaAla.....LeuG 690

```



```

2344 ACTGAAGTCTCGCGG..... 2358
597 eLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleA 614
2359 .....CGGAGCTGGCAGAGGCCATGGCTGCCACACACAG 2395
614 spThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
2396 CTGAGAGTGAAGTGTGAGCAGCTCGTCAAGAAGTAGCTGCCTGGCGTGAC 2445
631 GlyThrAlaAspGluAspGluGlyValProThrAspGlnAlaIleAr 647
2446 GGGTATGAGCATAGCCAGCAAGAGGAGGCACAGTATGGGCCCATGTTCCA 2495
647 gAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerP 664
2496 GGAACAGCTGATGACTTGAAG..... 2517
664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
2518 .....GAGGAATGTGAGAAGCCCGCCAGGAGCTGCAGGAGGCA 2556
681 AspGluLysValAla.....GluAlaAlaLeuGlnIlePheLys 693
2557 AAGGAGAAGGTGGCAGCATAGATCCACACAGCGAGCTCCAGATAAGCCG 2606
693 sAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaL 710
2607 GCAGCAG...AACAACTAGCTGAGCTCCATGCCAACCTGCCAGAGAC 2653
710 euLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
2654 TCACAGCAGGTC..... 2664
727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrG1 743
2665 .....CAAGAGAAGGAAGTCAG 2681
743 nPheAlaGlnIlePheGluProLeuHis..... 752
2682 GGCCCAAGAGCTTCAGATGACCTCTCCACTCTGCAGAGAAAGATGCGTG 2731
753 ..LysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuVal 768
2732 CCACCACAAGAGGTGGCCGCTTGAGACCTTGTC..... 2769
769 ThrIleGlyHisIleAlaLeuAlaProAspGlnPheAlaAlaProTr 785
2770 .....CGCAAGGCAGGTGAGCA 2786
785 pLysSerTrpValAlaThrPheIleValLysAsp...LeuLeuMetAsnA 801
2787 GCAGGAACAGCCCTCCGGGAGTAGTCAAGAGGCCTGCGAGGCGCAGGAG 2836
801 spArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGlu 817
2837 ACAGACAGCCCGAG.....TGGCTG.....GAAGAG 2862
818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValAr 834
2863 CAACAGGAGCCAGTCTCGCAGCACACAGCAGCG..... 2898
834 gTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrL 851
2899 ....CTCAGGCTATGGAGGGGAGGAGGAGCAGATGGGCAATGAGCTGG 2944
851 euArgLeuLeuThrIleLeuHisSerAspGlyAspLeuThrGluGln 867
2945 AACGGCTGGGGCCGCTGATGGAGAGCCAGGGGAGCAGCAGGAGGAG 2994
868 GlyLysIleSerLysProAspMetSerArgLeuAlaIleLys 884

```

```

2995 CQTGGCAGCAGAGGAGGAGGTGGCGCGCTGACCCAGGAGCGGGCGG 3044
884 rAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleThrL 901
3045 TGGCCAGGCTGACCTTGCCCTGGAG..... 3069
901 euGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917
3069 ..... 3069
918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuAr 934
3069 ..... 3069
934 gLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProv 951
3069 ..... 3069
951 allLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsn 967
3070 ..AAGCGGCCAGCAGAGCTTGAGATCGGCTGCAGAACGCCCTCAAC 3117
968 ValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLe 984
3118 GAGCAGCGT.....GTGGAGTTCGCTACCTGCAAGAGCACTGGC 3158
984 uSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
3159 TCATGCCCTGACGGAAGGAA.....GGCA 3184
1001 isAspProAspTyrValLysValGlnAspIleGlu.....GlnLeuLys 1015
3185 AGACACAGGAGTTGCCAAGCTTCGTGGTCTGAGGCGCCAGATAAAA 3234
1016 AspValLysGluCysLeuTrpPheVal.....LeuGluIleLeuMetal 1030
3235 GAGCTGGAGGAAGCTTCGCCAACCGTGAAGCAACTGAAGGAACAGCTGGC 3284
1030 aLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnI 1047
3285 TAAGAAAGAAAGGAGCAGCCATCTGGCTCAGGAGCCCAATCTGAGGCTG 3334
1047 leLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGlu 1063
3335 CTGCGCAGGACAGAG.....CCAACAGGCCCCCAAGCTGGAAGCA 3372
1064 LysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080
3373 CTGCGGCGCAGAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGA 3422
1080 rThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgp 1097
3423 GCAGGCTGACAGCTGGACGCGCCCTCGAGGCTGAGCGGGGCTCCCGG.. 3471
1097 hePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113
3472 .....GCTGAGCGGACAGTCTCTGGAGACTCTGCAGGCCCATTTAGAG 3516
1114 ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnVa 1130
3517 GAGAAGCCCGCAGGAG..... 3531
1130 lLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnT 1147
3532 .CTAGGCGCAGCTGAGTGCCTTAGCTCGGCCCAACGGGAGTTGGCTG 3580
1147 hr..LysSerSerArgMetGluThrValSerAsnAlaSerSerSerAs 1163
3581 CCITCCGCCCAAGGTACAGACACACAGAGGCTGAAGATGAGTGAAG 3630
1163 nProSerSerProGly.....ArgIleLysGlyArgLeuAspSerS 1177
3631 GCCCAGGTGGCGGGCGGCCAAGAGGCTGAGAGGAAAAAATAGCCTCAT 3680

```


55 LeuTyrLeuAsnLeu.....AlaLeuHisLeuAlaSerAspPhePh 68
||| :|:|:|:|:| ||| | | | | |
967 CTTTCTTTAAGCTGGGAGTTTGCACATCATCTG..... 1002
68 eLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuA 85
1002 1002
85 laAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAsp 101
1003CAG 1005
102 LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGl 118
:|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1006 CAGCTACAGGATGCCCTCAATGAGCTGACGGAG.....GA 1040
118 uAspThrLysSerProGln.....PheAsnArgTyrPheTyrLeuLeuG 133
| :|:|:|:|:|:| ||| :|:|:|:|:|:| :|:|:|:|:|
1041 GCACAGCAAGGCCACTCAGGAGTGGCTAGAGAAAGCAGGCCAGCTGGAGA 1090
133 luAsnIleAlaIleAlaIleValLysSerTyrAsnIleCysPheGluLeuGluAsp 149
:|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1091 AGGAGCTCAGCCAGCCCTGCAGGACAAGAATGC.....CTTGAAGAG 1134
150 SerAsnGluIle.....PheThrGlnLeuTyrArgThrLeuPh 162
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1135 AAGAACGAATCCTTCAGGGAATACTTTCACAGCTGGAAGACACTTGTGC 1184
162 eSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValA 179
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1185 CCAGCTGAGGATACCCACCCAGGAGAGAGGGCGAGGTG..... 1224
179 spLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeu 195
1225CTGGGTGATGCTCTGCAGCTGGAACCC 1251
196 LeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLys 212
||| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1252 TTGAAGCAAGAGGAGCCACTCTTGTGCAAAACAACACACACTCCAAGC 1301
212 sgAlaIleTyrAspLeu.....AlaLysAlaL 221
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1302 CAGGTGATGATGCTGGAGACTGCGGAGGCCAGCAGGAGCCAGCTGC 1351
221 euLeuLysArg.....ThrAlaGlnAlaIleGluProTyr 232
|| :|:|:|:|:|:| ||| :|:|:|:|:|
1352 TTGCTGAGCGGGCCACTTCGAAGAAGAAAGACAGCAGCTGTCTAGCCTG 1401
233 IleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSe 249
| | | | | :|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1402 ATCACTGACCTG.....CAGAGCTCCATCTC 1427
249 rAspLeuSerGlu.....HisV 255
| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
1428 CAACCTCAGCCAGCCAGGAGAGAGCTGGAGAGCCCTCCAGGCTCATG 1477
255 aIlePheAspIleLeuGluLeuTyrAsnIleAspSerHisLeu..... 269
||| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1478 GGGCCCGGTTGACTGCCCGGCTCTCTGACCTCTGAGCTCACCACA 1527
270 LeuLeuSerValLeuProGlnLeuGluPheLys.....LeuLysSe 283
||| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1528 CTCATATGCCACCATCCAGCAACAGAGATCAAGAACTGGCTGGCTGAAGCA 1577
283 rAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetP 300
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1578 GCAGGCCAAAGAGAGAGAGCCGCCAGCTAGCACAGACCCTC..... 1617
300 heGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1618CAACAGCAAGAACAGGCTCCAGGGC..... 1644

317 CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCy 333
||| :|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1645CTCCGCCACCAGGT 1658
333 sValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysA 350
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1659 GGAGCAGCTAAGCAGTAGCTGAAGCAGAGGAGCAGCAGTTG...AAGG 1705
350 spLeuThrGluTyrLeuLysValArgSerHisAspProGluAlaIle 366
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1706 AGGTAGCGGAG.....AAGCAGGAGGCAACT 1731
367 ArgHisAspValIleValSerIleValThr..... 376
| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1732 AGGCAGGACCATGCCAGCACTGCCCACTGCTGCAGAGAGCAGAGGC 1781
376 376
1782 CTCCTTAGGGAGCGGATGCGGTCTCAAGCAGCTGGAGCAGCTGGAGA 1831
377AlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeu 390
| | | | | :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1832 AGGAGAAGGCTCCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGCT 1881
391 AsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGl 407
||| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1882 AATGAAGCCCGGACAGTGCACAGACCTCAGTGACACAGGCCAGCGGA 1931
407 uAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr...AlaLeuGlnS 423
| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1932 GAAG.....GCAGAGCTGAGCGGAAGGTGGAGGAACCTCCAGG 1969
423 erAlaAla.....GlyLysAspAlaAlaLysGlnIle 433
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
1970 CTTGTTGTCAGACAGCCCGCAGGAACACATGAGGCCCGCCAGGT 2019
434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
||| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
2020 GCAGAGCTAGAGTTGCTGCGGTGTGAGCAGCAAAAAGCAACTGAGAA 2069
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467
:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:| :|:|:|:|:|
2070 AGAAAGGTGGCCAGGAGAGGAGCAGCTCCAGGAGCAGCTCCAGGCC 2119
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
|||:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2120 TCAAGAGTCTCTGAAGGTCAACAAG.....GGCAGC 2151
484 LeuAsp.....LeuAsnAlaValLysAlaLeuAsnGluMetTrpLys 497
|||:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2152 CTTGAAGAGGAGAGCGCAGGCTGCAGATGCCCTGGAAGAGCAGCAGCG 2201
497 sCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIle 514
|||:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2202 TTGTATCTCTGAGCTGAAGGAGAGACCCGAGCCCTGTTGGAGCAGCAT 2251
514 ysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
|||:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2252 AGCGGGAACGAAAGGAG..... 2268
531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
2268 2268
547 sLysPheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuG 564
|||:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2269CTGGAAGAGAGAGGCGCTGGGCCAAGAGGGGCTGG 2302
564 luValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
|||:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|:|
2303 AGGCTCATTTACTGCAGCTTGGGAGGCCCATCAGCTGAG..... 2343
581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPh 597

XX DT XX DE XX KW

AAQ54841;

```

alignment_scores:
  Quality: 179.50      Length: 1514
  Ratio: 0.377        Gaps: 60
  Percent Similarity: 42.801  Percent Identity: 18.428

alignment_block:
  US-09-512-581-2 x AAQ54841 ..

Align seg 1/1 to: AAQ54841 from: 1 to: 6306

16 ProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetVa 32
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
823 CCCAAGGAGCTTGAGGAGCTGCGTGACAAG.....AATGAGAGCCTTAC 866

32 IargArgLeuLysMetValValLysThrPheMetAspMet..... 45
   : ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
867 CATGGCGGCTGCATGAACCCCTTGAAAGCAGTGCAGGACCTGAAGACAGAGA 916

46 .....AspGlnAspSerGluGluGluLysGlu 54
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
917 AGAGCCAGATGGATCGCAAAATCAACACCGATTTTCGGAGGAGAAATCGAGAC 966

```


398 uAspLysArgTfPrpValArgLysGluAlaMetMetClyLeuAlaGlnI 415
3553 : : |||||
415 leTyLysLysTyrrAlaLeuGlnSerAlaAGlGlyLysAspAlaAlaLys 431
3569 TT.....AGTATTGTGGCAACAGAGCTGGG.....CTTAAG 3603
432 GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrtYrGlnAnSe 448
3604 GAAGTTCGATACAGGCCCTTAGATGTTGGCATCTGTTAGTTCGCCACT 3653
448 rIleAspAspArgLeu.....LeuValGluArgIlePheAla..... 460
3654 AAATGATGAAGAAGCTCTACTGTTATAGATAGCTAATATGCATTCATC 3703
461GlnTyrMetValProHis 466
3704 TTCAAAGTGCTCAGTTCACGGCAAACCTCCGAATAATAGTCTACAGT 3753
467 AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu..... 480
3754 ATACTGGATACCTTAATCAAAGAGAACCGCACCTGATTTTGAATAATAA 3803
481TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetT 496
3804 ACCTTACACTACTTGGCTTTAGTAGGCAAGCCCTGAATATAGGTATTTT 3853
496 rPLysCysGlnAsnLeuLeuArg.....HisGlnValLysAspLeuLeu 510
3854 CTCGCTGATGGCCAATTTCCAGAGATGGTGAATAAATAAGAGTACCAG 3903
511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSe 527
3904 GACCTTATA.....CCCATTTTCTTAATAACTTGAAAGT.....AG 3941
527 rLysValMetValIleThrArgAsnLeuProAsp..... 538
3942 TAAACAAGTATGTCATAAACCAAAATTTAGACGATATAGAGTATATCTC 3991
539Progly 540
3992 GGAGAAAGCAGACAGAAAGATCGATTGATTTTACACCAAGAGTTGGG 4041
541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGL 557
4042 CAAACTTCGTGATATACATATTAGTTTGGTGCTTTATTAGACACATCTC 4091
557 ulysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysL 574
4092 TAAGTTTTAGAAATTTAGACAGAGGACCTATGCGAGAAGTGGCCCAATGT 4141
574 ysGlnAlaClugLysValArgGluIleThrLysLys..... 586
4142 TCAGTATGATTGGT...GTTTATAGCGTTACAAGACGATGAGTTTAAAGA 4188
587LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMe 600
4189 ACAACATATTCAGAAACCAAGATTATGATTGATGATGATGATGATCAAC 4238
600 tIleLysPheLeuLeu.....GluArgIleAlaproValHisI 613
4239 TATTAAAGTTCTTGATATGGGTCATAAATGATATCTCGTCTCGCGTTT 4288
613 leAspThrCluSerIleSer.....AlaLeuIleLysGln 624
4289 GCACAAAGTGAGAAATCCCAGCAAGCAATTTGTTGTTGCCCTTGTCATAC 4338
625 ValAsnLysSerIleAspGlyThrAlaAspAsp.....G1 636
4339 GAATCATTAATAATATGGCGCTAAGTTTTCAGAGTCATGGGATATGAACCA 4388

[illegible]

necessary for G2/M cell cycle checkpoint control. This sequence was identified using the method of the invention for isolating a human checkpoint cDNA that is capable of restoring growth at a restrictive temperature in a yeast test cell, where the yeast test cell comprises a genome having a first gene that forms a DNA strand break at a restrictive temperature and a second gene that fails to induce a cell cycle arrest in response to the DNA strand break, whereby the growth of the yeast test cell is inhibited at the restrictive temperature, comprising:

(a) obtaining a human cDNA library comprising several human cDNA clones; (b) inserting the human cDNA clones individually into plasmid vectors containing a selectable marker gene; (c) transforming a culture of the yeast test cells with the plasmid vectors from the preceding step; (d) selecting for yeast test cells transformed with the selectable marker gene; (e) growing the selected transformants at the restrictive temperature and isolating a candidate transformant capable of growing at the restrictive temperature; and (f) identifying the human cDNA carried by the candidate transformant as a human checkpoint cDNA by sequencing the human cDNA carried by the candidate transformant and determining that the human cDNA is less than 50% homologous with both the first gene and the second gene. The nucleotide sequence of hURADcompB can be used to confer radiation resistance on a cell.

Sequence 8351 BP; 2835 A; 1522 C; 1557 G; 2437 T; 0 other;

alignment_scores:

Quality: 180.00 Length: 1348
Ratio: 0.321 Gaps: 68
Percent Similarity: 41.617 Percent Identity: 19.288

alignment_block:

us-09-512-581-2 x AAX01274 ..

Align seg 1/1 to: AAX01274 from: 1 to: 8351

```

68  PheLeuLysHisProGlyLysAsp.....ValArgLeuLeuValAlaCy 82
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2359 TTCTATCTCACCAATATAGTAAATTTAGCGAATCTTTGTTAAGTG 2408

82  sCysLeuAlaAspPheArgIleThrAlaProGluAlaProTyrThrS 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2409 CATCTTATTTCTCTACATAGGATATTTTCA.....CACTTTT 2446

99  erProAspLysLeuLysAsp..... 105
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2447 AACCTCCAAACCTTACAGATGGAACGGTCAATCAATAAGACCTTAA 2496

106 .....IlePheMetPheIleThrArgGlnLeuLys..... 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2497 CTGGTACAAAGTCTTTATGAATTTCTAACAGATACCTACGTTTATTAA 2546

116 .....GlyLeuGluAspThrLysSer.ProGln 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2547 TACTAGAAATATACCTTTATTCATATATACAGACTCTCATATTCGGA 2596

125 PheAsnArgTyr.....PheTyrLeuLeuGluAsnIleAl 136
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2597 ATGACACACTGCCGCTCATAAAGTTTCTACAATCTCAAAAATTGCCA 2646

136 atpValLys.SerTyrAsnIleCysPheGlu.....LeuGluAsp 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2647 GTGGTGAAGAAATTTAGTCATTTGCTGGACACAAATTAACATTTGAC 2696

150 .SerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIle 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2697 TTCTAATGATGATTTGATACACTACTTTTGAACACTGATTTGATTTTCA 2746

166 snAsn..... 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2747 ATTCTGATGATTATAGTTTACGATAATGATGACTTTTGCAAAATTA 2796

167 ..... 167

```

```

2797 ATGGCCAAAATTTTAAAGAAAACACACCATATCAATTAATCTATCGCCTATTTT 2846
168 .....GlyHisAsn.....GlnLysVal. 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2847 ACCTGTATTACTAAGACAGCTTGGAAAAAACCTCGTGGAAAGAAAGTTG 2896
174 ..HisMetHisMetValAspLeuMet..... 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2897 GCTTTCAAAATTAATAGAATATTGGGATATCTTCAAAAACAATTTCTC 2946
181 ..... 181
2947 GATATTTTCCAGAGATATATCAATCCCTTATGCAATATTCAATATAAGAG 2996
182 .....SerSerIleIleCysGluGlyAspThrValS 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2997 CGATGTGCTAAGTGAATTTGCTAAGATTTATGTGTATGGCGATACA.... 3042
192 erGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLys 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3043 .....AGTTTAATTAACCAAAATGAAGTTTAATTTACTG.....AAA 3078
209 AsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgTh 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3079 AAAACACAGTAGGCAATATTT.....GCCGTAGCTTTGGTAAACACGG 3122
225 rAlaGlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuM 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3123 ATTATTTTCTTTGGATATCTTGGAAACCCCTTTTAAATATAGGCGTCCAA 3172
242 etLeuGlyLysThrSerIleSer.....AspLeu 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3173 CTTTTCACAAAGGATATATACTGCATACCTCCCGATATTATAAACTTTA 3222
252 SerGluHisValPheAspLeuIleLeuGluLeuTyrAsn..... 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3223 GCTGAA.....ATAACGAAAGCTCTACAAAAAACACGCGTTAC 3257
265 .....IleAspSerHisLeuLeuLeuSerValLeuProGlnLeuG 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3258 TAAAGATGAAGTGACGAGAGATGCTAATATGATTTATGCTCTTTGCG 3307
278 luPheLysLysSerAsnAspAsnGluGluArgLeuGlnValValLys 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3308 GATTTTAAATCACCAATTTTGAAGAAAGACAAAGG..... 3342
295 LeuLeuAlaLysMetPheGlyAlaLys..... 303
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3343 .....CATGGTTGAAAGTACAAAAAATATCAATACTGGAC 3377
304 .....AspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCystyrL 319
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3378 GGATGATCAGGAACAAGCGTTCCAAAGAAA...CTACAGGATAAATATCT 3424
319 euGlyArgPhe.....AsnAspIleHisValProIleArgLeu 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3425 TAGGTATTTTCCAAGTTTTCGAGTGACATACAT..... 3459
332 GluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAl 348
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3460 .....GATGTGA 3467
348 aLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluA 365
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3468 AGGCCGCCACCTTACTACGAAAAGTTAAG..... 3498
365 laIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAsp 381
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3499 ..GTTATCAATGCGATTTCTTTCTTATCATATATGCCCCCAAAAAATCA 3546
382 IleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLe 398
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3547 ATAATT..... 3552

```



```

4874 AACAAATTTACAGTAACACTACACGCGACTTTTATCATTAAGATACGAAA 4923
757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773
4924 ACAACTAACATG..... 4935
773 eAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpVal 790
4936 ....CTTTTGAGAAATAGATGAGTCTTTTGGGAAACAACCCCTTCA..... 4974
790 laThrPheIleValLysAspLeuMetAsnArgLeuPro..... 804
4975 .....GATTGCTAGCTCAACGCTCTAGAGCGGAT 5007
804 ..... 804
5008 TCTTTTGAAGGTCGCTCTATACCTTGAACAGTGTATCGACAGATCC 5057
805 .....GlyLysLysThrThrLysLeuTrpValProA 815
5058 TCACGATAAGAACCAAAATGGCAACTACTGAAAAATTTACAAATCACAT 5107
815 spGluGluVal..... 818
5108 ACGAAGAAATAGGAGACATGACTCAGTCGATCGTACTGAGAACCTTT 5157
819 ...SerProGluThrMetValLysIleGlnAlaIleLysMetMetValAr 834
5158 GCTACAGGAAATCGTTTCTTAAATTTGAAGTTCGCAATATCTCGAAAA 5207
834 gTrp.....LeuLeuGlyMetLysAsnAsnH 843
5208 CTGGAACCTCGCACAGACTGCTTTAATCTCTCGCAAAATTTTCAGATG 5257
843 isSerLysSerGlyThrSerThrLeuArg.....Leu 853
5258 ACCCCAAACTACACAGAGTCTGCTTAAAGTCTATGATGACCACCAATTTG 5307
854 LeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysI 870
5308 TATTCTCAATAATATTCGAACTCTCGTTCATTTTCAGCGGAAAAAT 5357
870 eSer...LysProAspMetSerArgLeuArgLeuAlaGlySerAlaI 886
5358 TTTCTTTGTCCTCCAGATGTGAAG..... 5379
886 leValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGlu 902
5380 .....GAA 5382
903 GlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGlu...CysTyrGlnVa 918
5383 TGGTACAGCATAGGCTTCAAGCTGCAATCTAGAAGCAATGTTCAAC 5432
918 lArgGlnValPheAlaGlnLysLeuHis.....LysGlyLeuSerArgL 933
5433 TTTGAAATTTGGTAGAACAAATAGAGAGTTTAAAGAAATATTGACGATA 5482
933 euArgLeuProLeuGluTyr.....MetAlaIleCysAlaLeuCys 946
5483 GAGAAGTACTTTTGCAGTACAATATTGCGAAAGCTTAAATGCCATCTCA 5532
947 AlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuVa 963
5533 AACGAGATCCATTAGGACTCAA..... 5556
963 lLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValS 980
5557 .AAATACATCCACAATTCCTTTAGGCTTATCGGAACAAATTTTAAACGT 5605
980 erGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996

```

```

5606 CATCTAAAGAGAGACGCGCTGCTAAAGAAACAGAAATTTATTGATGAATTA 5655
997 HisLeuLeuAlaHisAspProAspTyrVal...LysValGlnAspIleG 1012
5656 CACAGTTTA...TATGACCTCAGTTTTTTTATCTTCTGCGAAGATAGTT 5702
1012 uGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeu 1029
5703 TGAATACAAAGTAACACTACCATCTCGATTATCGAATGAACGATTG 5752
1029 eAlaLysAsnGluAsnAsnSerHisAlaPhe...IleArgLysMetVal 1044
5753 GGGCTGACTCTGTCACAAATCATATCATATTGTCAATGAGAAAGTCATT 5802
1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMe 1061
5803 GACCAATTGAAATGAATGACACAGCA.....GACGCTGACTT 5840
1061 tAsnGluLysLeuTyrThrValCysAspValAlaMetAsn..... 1074
5841 AGGAAAAACATTTCTTACCTTTAGCCCAATTTGCGAGAAACAACGCTAGGC 5890
1075 ...IleIleMetSerLysSerThrThrTyrSerLeuGlu 1086
5891 TAGATATAGCCTCCGAATCATTAATGCTGTTTGGAA 5928
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AA01274
seq_documentation_block:
ID AAX01274 standard; cDNA to mRNA; 8351 BP.
XX AC AAX01274;
XX XX AC
XX DT 08-APR-1999 (first entry)
XX DE Yeast ME1 coding sequence.
XX KE Cell cycle checkpoint gene; yeast; radiation resistance; detection;
XX KW G2/M checkpoint; ss.
XX OS Saccharomyces cerevisiae.
XX FH Key Location/Qualifiers
XX FT CDS 784..7890
XX FT /*tag= a
XX XX US5866338-A.
XX PN
XX PD 02-FEB-1999.
XX XX
XX PF 06-JUN-1997; 97US-0870693.
XX XX
XX PR 18-FEB-1994; 94US-0198446.
XX PR 12-MAY-1992; 92US-0882051.
XX PR 14-MAY-1992; 92US-0884426.
XX PR 12-MAY-1993; 93WO-US04458.
XX PR 06-JUN-1997; 97US-0870693.
XX XX
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PA (UNIV-) UNIV ARIZONA.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Groudine MT, Hartwell LH, Plon SE, Weinert TA;
XX XX WPI: 1999-141919/12.
XX DR P-PSDB; AAW73893.
XX XX
XX PT Nucleotide sequences that hybridise to huRAD cDNA - method for
XX XX isolating human checkpoint cDNA, and antibody to CDC34
XX PS Example 4; Column 61-78; 73pp; English.
XX XX
XX CC This sequence encodes the yeast ME1 protein. The DNA sequence is

```


[illegible]


```

971-u.....TyrLeuLysGlnHisA 977
:          |||||
2123 AGAATGGGTCCTGGGAGAGAGAGTTTGGAGAGATATTAAACAGCCAT 2172
977 laalavalserGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993
:      |||||
2173 CCACAGAGTGTACCTATTTCTATCTAGCCCTTGGCAACGTGTGGCTC 2222
994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
: |||||
2223 CAACCTTTACATCAGCCCGGAGATCGAGAAAGAAAGAAAGCGTCTCA 2272
1010 pilleGluGlnLeuLysAspValLysGluLysLeuTyrPheValLeuGluI 1027
: |||||
2273 AGATCGTGTCTGCGCATCTACAAACAGTACTCAGAAATGATGCAAGA 2322
1027 leuMetAlaLysAsnGluAsnAsnSer.....HisAlaPhe 1039
: |||||
2323 ATCTGTATGCTGCCAATGGCATAGGAGCTGTTTGGCCCAACAAAGGATAT 2372
1040 IleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPr 1056
: |||||
2373 TTTCGTAAGCTCGTGTATTTGCCCAAGTAAGAGAACACAGCAGA 2422
1056 o...AspAlaLysMetAsn..... 1062
: |||||
2423 TATTAGTGTGTGCTGACTGACCTAGCACACATCTATGTGGACAAAGC 2472
1063 .....GluLysLeuTyrThrValCys..... 1069
: |||||
2473 AGTACATCAGCGCGCTGCATGTATGAAACCTGCCTCCGAAAGTTCTAT 2522
1070 .....AspValAlaMetAsnIleMetSerLysSerTh 1081
: |||||
2523 AGCACCAAACTGAAAGTTGACTCTATTGGCCCGGCCCTCTTCAA 2572
1081 rThrTyrSerLeuSerProLysAspProValLeuProAlaArgPhep 1098
: |||||
2573 GTGTGGCAAGTTACAGGAATGCAACAGACTTTGCTCAAGGCTAGACATG 2622
1098 heThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
: |||||
2623 TGGCACCCAGTATACAGTTCTTATGTTTAATGTGCCTTGGTCCGCAA 2672
1115 GluMet...LysSerPheThrProGlyLysProLysThrThrAsnVa 1130
: |||||
2673 AGATTAGCTACCTCTGCTGGAAGATGAAAGAAAGTAATCTGAAGGAAGT 2722
1130 leuGlyAlaValAsnLys.....ProL 1138
: |||||
2723 ACTTAATGCTGTGAAGAACTGGAGCTTGCACATAGACTACTTCAGTTATT 2772
1138 euSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThr 1154
: |||||
2773 TGAGTAAAGTGGAGATAAATGAGATTTGATTGGCCCTTGCTGTACA 2822
1155 ValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLy 1171
: |||||
2823 GAAGCCAGGCGAGTGTCTGACTTACTGAGCCAGGCCAGTACCATTGTG.. 2870
1171 sGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrT 1188
: |||||
2871 .....GCCCGGCACGCAACAAAGATGAAGAGCGGGAG..... 2906
1188 hrMetSerSerProLeuProGlyLysLysSerAspLysArgAspSer 1204
: |||||
2907 .....CTCGGGCCCAAGCAAGAGCAAGAAAG 2933
1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrPr 1221
: |||||
2934 GAGCTGTTAAGGCAGAACTCTTAAGAACAGGAGAGAAACGCTCTCAG 2983
1221 oValThrGluGlnGluLysLeu..... 1229

```

```

|||
2984 AGAAAAGGAGACGACAAAAGAAACTTTTGGACAGCGGCCCGCCAGTATGTGG 3033
1230 .....GlyMetaspLeuThrLys 1236
: |||
3034 AGAAGACCAAAATATTTCTATCTTTACTGGTGAGACTGAAGCAACAAAA 3083
1237 LeuValGlnGlnLysProLysLysSerGlnArgSerArgLysArgL 1253
: |||
3084 ...GAGAAGAAAAGAGGTGGTGGTGGAGCGGCTTCTAAGAAGGGAGG 3130
1253 yHis..... 1254
: |||
3131 AGATTGATGAATTTGTCATGATGACACTGATGATGACCTACCTATAT 3180
1255 .....ThrAlaSerGluSerAsp 1260
: |||
3181 CAAAAGAAAGAGAGAGAAAGGTAGTGGCAGTGAACAAGAGGTGAA 3230
1261 GluGlnGlnTrpProGluGluLysArgLeuLys..... 1271
: |||
3231 GATGAGGAGGTGGTGAGAGAAAGAAAGAGAGAGAGAGACATCCAAA 3280
1272 ...GluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysL 1287
: |||
3281 GGGAGAAGAGGATCTCATGATGATGAACAGAAATGGCCCC.....A 3324
1287 ysglyLysArgGlyArgProProLysProLeuGlyGlyThrProLys 1303
: |||
3325 AACCAAAAAACGACGTCCACCAAAAGCAGAGAGAGAAAGAGGTCCTCAAG 3374
1303 ..... 1303
3375 CCAGAACGTGCTGCTCCATCAATGAAGGGGAAATAAATCCAAAGCCAT 3424
1304 .....GluGluProThrMetLysThrSerL 1312
: |||
3425 AATTTCATCAAGTATGACTCTTCGGATGAGGATAAACTTAAATTCGTG 3474
1312 yslLysGlySerLysLysSerGlyProProAlaProGluGluGlu 1328
: |||
3475 ATGAAGGACATCCCAAGACAGCAACAGACAGTGTACTCAGACGAGGAC 3524
1329 GluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysG 1345
: |||
3525 GAACACAGAAAGAAATGTGCTCATCAGAGATGATTCCTGATGAGAACCA 3574
1345 nHisArgValSer.....ArgA 1351
: |||
3575 GAACAAGTCTGGCAGCGGCGGCGAGTCCCGGAGGCCAGCAAGACAGC 3624
1351 rgAlaGlnGlnArgAlaGluSerProGlu..... 1360
: |||
3625 GGTGATCAGACTCAGACAGTACCAAGCAGCAGCAGCAGCAAGAGAGGCC 3674
1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyAr 1377
: |||
3675 TCCGGTCTGAGCAGTCTGACAATGAATCTGTGCGATCAGGAGAGGCCA 3724
1377 gProSerLysThrProSerProSerGlnPro 1387
: |||
3725 CTCAGGAGTTTCTGAGAACGACTCTGCCCA 3755
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT91039
seq_documentation_block:
ID AAT91039 standard; cDNA to mRNA; 8351 BP.
XX
AC AAT91039;
XX
DT 25-FEB-1998 (first entry)
XX
DE Yeast checkpoint control gene MEC1 cDNA.

```



```

1189 .....MetSerSerProLeuProGlyLysLys..... 1197
728 GGCCTGGCCGACGCGAGGCTCTGGAGGACGCAAGAAGAAGG 777
1198 .....SerAspLysArgAspSerAspLeuValArg 1208
778 TGCCATCAGCCTCGGATTCAGACTCCAAAGCCGACTCAGAT..... 818
1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluL 1225
819 .....GGGGCAAGGAGGACCTGTGTTACAGC 847
1225 nGluGluLysLeuGlyMetAspLysLeuValGlnGluL 1242
848 ACAG..... 851
1242 ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258
852 ..CCGTCTCATCTCTCTGCTGCTGCTCTCTCATCATCTCTCTCAGAC 899
1259 SerAspGluGlnTrpProGluGluLysArgLysGluAspIleLe 1275
900 TCAGATGTCT..... 911
1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyA 1292
912 .....GTCAAGAAACCCCTCGAGGTA 933
1292 rg.....ProProLysProLeuGlyGlyThrPro 1302
934 GAAAGCCAGCTGAGAGCCACCCCAAAACCC...CGAGGGCGGAGACCA 980
1303 LysGluGluProThrMetLysThrSerLysLysGlySerLysLysSe 1319
981 AAGCCAGACGACCACTCTCCACCTCCAGCTGACAGTGCAGTGACAG 1030
1319 rGlyPro..... 1321
1031 TGGTGAAGTAGACCGCATCAGTCAATGGAAGACGCTGATGAAGACGAA 1080
1322 .....ProAlaProGluGluGluGluGluGluGluGluGlnSer 1334
1081 GCGGTGAGCTGGAGGCTCGAGCGCAGCGGAGGAGGAGGAGCTCCGG 1130
1335 GlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArg 1351
1131 AGACTTCGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1169
1351 gAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrG 1368
1170 .....AAGGCGGCGAGAGCGGTGGGGGCGACGAGTGGAGAGGAGCTGGAG 1215
1368 lnSerThrPro.....GlnLysGlyArgGlyArgProSer 1379
1216 ACGAGGAGCTGTGAAGAAGCGTAGCGAAGGCTCGAGGCGCGAGGCACA 1265
1380 LysThrProSerProSerGlnProLys 1388
1266 CCATCTCTCTGACTCAGAACCTGAG 1292

```

seq_name: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ86236

seq_documentation_block:

ID AAQ86236 standard; cDNA to mRNA; 4548 BP.

AC AAQ86236;

DE 30-NOV-1995 (first entry)

DE Megakaryocyte stem cell maturation factor clone 1204B.

KW Megakaryocyte stem cell maturation factor; clone 1204B; diagnosis; treatment; abnormal differentiation; megakaryocytic leukaemia; ss.

```

XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 633..3908
XX /*tag= a
XX
XX JP07067658-A.
XX
XX 14-MAR-1995.
XX
XX 03-SEP-1993; 93JP-0243767.
XX
XX 03-SEP-1993; 93JP-0243767.
XX
XX (MOCH ) MOCHIDA PHARM CO LTD.
XX
XX WPI; 1995-143852/19.
XX
XX P-PSDB; AAR70982.
XX
XX Novel DNA involved in differentiation of stem cells to
XX mega:karyocyte(s) - used in diagnosis and treatment of diseases
XX involving abnormal differentiation, e.g. megakaryocytic leukaemia
XX
XX Claim 1; Pages 18-23; 41pp; Japanese.
XX
XX AAQ86236 encodes AAR70982 the megakaryocyte (MKC) stem cell maturation
XX factor clone 1204B. The DNA is involved in the differentiation of
XX stem cells to MKCs and platelets, it can be used as a probe to
XX diagnose diseases involving abnormal differentiation, it may also
XX be used in the treatment of MKC leukaemia. Antibodies raised
XX against the protein can be used to immunologically measure the MKC's
XX level of differentiation.
XX
XX Sequence 4548 BP; 1431 A; 923 C; 1096 G; 1098 T; 0 other;

```

alignment_scores:

Quality: 180.00 Length: 1277
Ratio: 0.330 Gaps: 54
Percent Similarity: 42.756 Percent Identity: 18.168

alignment_block:

US-09-512-581-2 x AAQ86236 ..

Align seg 1/1 to: AAQ86236 from: 1 to: 4548

```

344 HisProAspLeuAlaLysAspLeuThrGluTyrrLeuLysValArgSerHi 360
||| ||| ..... ||| .....
402 CATCGAGATTCCTCCGGGACACTGACGAGGTCAATTGAACCTT..... 413
360 sAspProGluGluAlaIleArgHisAspValIleValSerIleValThrA 377
||| ..... ||| ..... ||| .....
444 .GACTTCGATCAGTTACCGGAGGAGAGTAAAGTTATCATCTATTCTGAAA. 491
377 laAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
491 ..... 491
394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe 410
||| ..... ||| .....
492 CAGGAACACACACAACTGCACATATGG.....AT 520
410 tGlyLeuAla...GlnIleTyrrLysLysTyrrAlaLeuGlnSerAlaAlaG 426
||| ..... ||| .....
521 TGCTTTGGCGCTGGAATACATAAGCAAGGAAAAACAGAGAGTTTGTAA 570
426 LysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHis 442
||| ..... ||| .....
571 AATCGTTGGAAGCAGCAGTATAGATGGCAATTTGGAC..... 608
443 IleTyrrTyrrGlnAsnSerIleAspArgLeuLeuValGluArgIlePh 459
||| ..... ||| .....

```


1385 rGlnProLysLysAsnVal 1391
 :::::::::::::::
 2189 AGAACCAAGCGCAGGTG 2207

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV01731

seq_documentation_block:

ID AAV01731 standard; cDNA; 2108 BP.

AC AAV01731;
 XX
 DT 27-MAR-1998 (first entry)
 XX
 DE Mouse liver cancer-originated culture cell growth factor encoding cDNA.
 XX
 KW Mouse; human; liver cancer-originated culture cell growth factor;
 KW HHDGF; HET-A; HET-B; ds.
 XX
 OS Mus sp.
 XX
 FH Location/Qualifiers
 FT 99..2108
 FT /*tag= a
 FT /product= "Liver cancer-originated culture cell growth
 FT factor"

XX JP09313185-A.

XX 09-DEC-1997.

XX 27-MAY-1996; 96JP-0131788.

XX 27-MAY-1996; 96JP-0131788.

XX (KISHI/) KISHIMOTO C.

XX (SEKI) SEKISUI CHEM IND CO LTD.

XX WPI; 1998-080076/08.

XX P-PSDB; AAW37483.

XX DNA segment encoding protein homologous to human liver

XX cancer-originated culture cell - which may be modified to produce

XX polypeptide of at least 5 continuous amino acids, useful in

XX producing, e.g. protein HET-A

XX Claim 9; Page 10-12; 18pp; Japanese.

XX The present sequence encodes a protein which has local homology to a
 CC human liver cancer-originated culture cell isolated from a mouse
 CC testicle cDNA library. The sequence may (1) be modified to produce a
 CC polypeptide comprising at least 5 continuous amino acids; (2) a
 CC polypeptide protein similar to protein of (1), but comprises no protein
 CC combined naturally and also has 2 amino acid sequences (as given in
 CC the specification) which may be modified to produce a sequence of at
 CC least 5 amino acids; (3) a recombinant DNA molecule comprising a vector
 CC and the DNA segment of (1); (4) a cell comprising (3); (5) producing a
 CC polypeptide comprising amino acids corresponding to a protein of (1),
 CC in which a cell comprising (3) is cultured under a conditions enabling
 CC the expression of (1) to produce the polypeptide which is isolated.
 CC The above method may be used to produce proteins HET-A and HET-B.

XX Sequence 2108 BP; 567 A; 557 C; 681 G; 302 T; 1 other;

alignment_scores:

Quality: 180.00 Length: 543
 Ratio: 0.800 Gaps: 23
 Percent Similarity: 41.436 Percent Identity: 20.994

alignment_block:

US-09-512-581-2 x AAV01731 ..

Align seg 1/1 to: AAV01731 from: 1 to: 2108

931 SerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAl 947
 ||| ::::::::::|
 20 TCAGCGCTCCGACTCCGCTCTGACTTCGCTCTCACCGCGTGTGTCG 69
 947 alyaspProValLysGluArgAlaHisAlaArgGlnCysLeuValL 964
 ||| ::::::::::|
 70 TTCCCGCGCGCTCCGCTCCGCGCAGCATCGCAGCGCTTCAAGCCC 119
 964 ySAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
 ::::::::::|
 120 GGGGACTTGGTGTTCCTCAAGAT. 142
 981 GluLysLeuLeuSerLeu. 986
 ||| ::|
 143 GAAGGCTACCCGCACTGCCGCCGGATTGATGACATTGCTGATGGTG 192
 987LeuProGluTyrValProTyrThrIleHisLeuLeu.Ala 1000
 ||| ::|
 193 CGGTGAAGCCGCCCAACAAATACCCCATCTTCTT...CTTGAACG 239
 1001 HisAspProAspTyrValLysValGlnAspIle. 1011
 ||| ::|
 240 CATGAACGCGCTTCTCGGACCCCAAGGACCTTCCCTCATCATAGTG 289
 1012GluGlnLeuLysAspValLysGluCysL 1021
 ||| ::|
 290 CAAGGACAAGTACGGAAGCCCAACAAGAGAAAGCTTCAATGAGGGC 339
 1021 eutrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnSerHis 1037
 ||| ::|
 340 TCTGGGAGATC.CAGAACAACCCCAT 365
 1038 AlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaG 1054
 ||| ::|
 366 GCC.....AGCTACAGTCGCGCTCCGCGGTGAGCTCCTCGACATGA 409
 1054 nGlyPro.....AspAspAlaLysM 1061
 ||| ::|
 410 GGCCCTGAAGCGCCTGGTGTGGCAGCGACGTGGACAGGACAAAG 459
 1061 etAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleMet 1077
 ||| ::|
 460 AATCCGCGAGGTGATGATGTG.....ACAGCTGTGACC 494
 1078 SerLysSerThrTyrSerLeuGluSerProLysAspProValLeuPr 1094
 ||| ::|
 495 ACCACAGCCACCATGACAGGATGGAGCGGATTCCTGAC..... 533
 1094 oAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnT 1111
 ||| ::|
 534TCTCACAAGAGCAGTATCAGT..... 557
 1111 yrLeuProGluMetLysSerPhePheThrProGlyLysProLysThr 1127
 ||| ::|
 558GGCTGAAGCGGAAG 572
 1128 ThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysG 1144
 ||| ::|
 573 ACACCAAGTCTTGAAGTG.....TCAGTCTCTAAACG 604
 1144 nSerGlnThrLysSerArgMetGluThrValSerAsnAlaSerSerS 1161
 ||| ::|
 605 AGCTAGAGAGAGCTTCCAGTCAGCTGGATCAGGCGCAGTGTGTCCTCATCAG 654
 1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSer 1177
 ||| ::|
 655 AAGAGGAGCTCTGAGGCCCA.....TCT 677
 1178 GluMetAspHisSerGluAsnGluAspTyrThr..... 1188
 ||| ::|
 678 GAGTCGGAGAAGACCAAGTCACCGAGGATTTACCCCAAGAGACAGCAGC 727

| | | | |
|--------------------------|---|--|------|
| | 1739 | GAAGCTTTTACCAGCACTGGCTTAATAAAAGCTTCTCGAAAACACAGA | 1780 |
| | 1287 | ysGlyLysArgGlyArGProProLeuGlyGlyGlyThr..... | 1301 |
| | 1789 | AGGGAAAAGAGGAGGAGGAGCAGGAATCAGGGGAGCACACTCAGGTT | 1838 |
| | 1302 | ProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLy | 1318 |
| | 1839 | CCAGCGATTCTCCGGACACGCCAGGAGGAGCAAAGGGCGAGAGCTCGC | 1888 |
| | 1318 | sSerGlyProProAlaProGluGlu..... | 1326 |
| | 1889 | CTCATCCCTGAGGAGCCGAGAGATCAGTGCTCTCGAAAAGGGCTTAG | 1938 |
| | 1327 | GluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys | 1340 |
| | 1939 | CCGAGGTGCAGCAGATGGGAAGCTGAAGAAGGAGCTACTTCCGATGGA | 1988 |
| | 1341 | SerLysSerLysGln..... | 1345 |
| | 1989 | GAGAAAAAAGAGAGAGGTGTCACTCCCTGGGCATTCATAAAGATGGT | 2038 |
| | 1346 | HisArgValSerArgArgAlaGln..... | 1353 |
| | 2039 | GACGCCCAAGAAGCGTGTAGACGGCTTCGGAAGAGTGATAAAGAGATG | 2088 |
| | 1354 | GlnArgAlaGluSerProGluSerSerAlalleGluSerThrGln | 1368 |
| | 2089 | AGCTGGACAAGGTCAAGAGCGTACTCTGTCTCCACCGAGAGCACAGCC | 2138 |
| | 1369 | SerThrProGlnLysGlyArgGlyArGProSerLysThrProSerProSe | 1385 |
| | 2139 | TCTGAATGCAAGAAGAAATGAAGGAGCGGTGAAGAGCCAAAGCCGA | 2188 |
| | 1385 | rGlnProLysLysAsnVal | 1391 |
| | 2189 | AGAACCAAGCGCAGAGTG | 2207 |
| seq_name: | /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA74903 | | |
| seq_documentation_block: | | | |
| ID | AAA74903 standard; cDNA; 6605 BP. | | |
| XX | AAA74903; | | |
| AC | | | |
| XX | 26-JAN-2001 (first entry) | | |
| DT | | | |
| XX | Human gravin coding sequence. | | |
| XX | | | |
| KW | Human; gravin; PKA RII binding site; myasthenia gravis; | | |
| KW | kinase anchoring protein; CAMP dependent protein kinase; ss | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | CDS | 192..5534 | |
| FT | | /*tag= a | |
| FT | | /product= "gravin" | |
| XX | | | |
| FN | USG090929-A. | | |
| XX | | | |
| PD | 18-JUL-2000. | | |
| XX | | | |
| Pf | 19-DEC-1997; 97US-0994570. | | |
| XX | | | |
| PR | 19-DEC-1996; 96US-0769309. | | |
| XX | (UYOR-) UNIV OREGON HEALTH SCI. | | |
| PA | | | |
| XX | | | |
| PI | Klauck TM, Scott JD, Nauert JB; | | |
| XX | | | |
| DR | WPI; 2000-523763/47. | | |


```
5291 ATCATTCAGGCTGACACACTTTTGGATGAATCAGAGAAAAGAACCCCA 5340
1114 OGU.....MetLysSerPhePheThrProGlyL 1124
5341 GCAAAAGAGAGAGTCTTAAGCGTTTAAAGGAGCAACTGATGACATAC 5390
1124 ySProlYsThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
5391 GCCAAGGTGGAGCTCTACACGTGACCAAGCAAACTGTGATGGCAAC 5440
1141 Alagly.....LysGlnSerGlnThrLysSe 1149
5441 CGCGGTGACCACTGCAGGAAATTAGTAGCCCAAAATCTCAGAGCTCAA 5490
1149 rSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerS 1166
5491 CCATCGATTTCAGCCATTTCACACAGAAATTAAAGACTGGAAGCCCTCA 5540
1166 erProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSer 1182
5541 TTCTCT.....TTGAAGGAATTGGAGCAGTTT 5566
1183 GluAsnGluAspTyrThrMetSerSerProLeuProGlyLys..... 1196
5567 AACTCAGATATACAAAATTGTCTGAACCACTGGAGGCTGAATTCAGCA 5616
1197 .....LysSerAspLysArgAspAspS 1204
5617 GGGGGTGAATCTCAAAGAGGAGCACTTCAATAAGATATGAATGAAGACA 5666
1204 erAspLeuValArgSerGluLeuGlyLysProArgGlyArgLysLysThr 1220
5667 ATGAGGCTACTGTAAGAATAATTGTCAAAGAGGAGGACAACTTACAACA 5716
1221 ProValThrGluGlnGluLysLeuGlyMetAspLeuThrLysLe 1237
5717 AGAATCACAGATCAGAGAGAGAGAGGAATAAAGATAAAGACAGCAGCT 5766
1237 uValGlnGluGln.....LysProLysGlySerGlnArgSerA 1250
5767 GTTACAGACAAAACATAATCTCTCAAGGATTTGAGCTCTCAAGAAGAA 5816
1250 rLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGlu 1266
5817 AAAG.....GCTCTAGAAATTTCTCATCAGTGTGATCAG 5851
1267 GluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSe 1283
5852 TACAAGAGGAGGAGCTGATGATCTCTCTGAAATGCTTGGATGACATT.... 5896
1283 rProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyG 1300
5896 ..... 5896
1300 lyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLys 1316
5897 .....GAAAA 5902
1317 LysLysSerGlyProAlaProGluGluGlu.....GluGluG 1330
5903 AAATTAGCCAGCTACCTAGGCCAGAGATGAAGGAATAAAGGAAT 5952
1330 uGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisA 1347
5953 TGATCGGGAATTG.....CAGAAAGAAAGAGAGGAGCTGAATG 5990
1347 rgValSerArgAlaGlnGlnAlaGluSerProGluSerSerAla 1363
5991 CAGTGCCTAGCAGCTGAGGGCTGTGTCAGGATGGCGCCGCAATGGCA 6040
1364 IleGluSerThrGln 1368
:::|||||
```

6041 GTGGAGCCAACTCAG 6055

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV23545

seq_documentation_block:

ID AAV23545 standard; cDNA; 6605 BP.

XX AAV23545;

XX 13-JUL-1998 (first entry)

XX Human gravin polypeptide coding sequence.

XX Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;

XX CAMP-dependent protein kinase; protein kinase C; autoimmune disease;

XX Myasthenia gravis; nicotinic acetylcholine receptor; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 192..5534

XX /*tag= a

XX US5741890-A.

XX 21-APR-1998.

XX 19-DEC-1996; 96US-0769309.

XX 19-DEC-1996; 96US-0769309.

XX (UYOR-) UNIV OREGON HEALTH SCI.

XX Klauck TM, Nauert JB, Scott JD;

XX WPI; 1998-260552/23.

XX P-PSDB; AAW53863.

XX New polypeptide fragments of protein kinase binding protein gravin -

XX are useful for the study of modulation of action between gravin and

XX protein kinase(s)

XX Example 1; Column 19-32; 32pp; English.

XX This sequence encodes the human gravin polypeptide, which represents

XX a polypeptide of the invention. The polypeptides are fragments capable of

XX binding to type II regulatory subunit of CAMP-dependent protein kinase

XX (PKA). Gravin is a kinase anchoring protein that binds to type II

XX regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an

XX antigen of the autoimmune disease Myasthenia gravis (MG), where a patient

XX develops antibodies against their own nicotinic acetylcholine receptors.

XX The polypeptides are useful for providing analogues of gravin in the

XX study of the modulation (e.g. blocking, inhibiting and stimulating) of

XX interactions between gravin and kinase. The peptides are involved in the

XX modulation of gravin-kinase interactions.

XX Sequence 6605 BP; 2091 A; 1416 C; 1826 G; 1272 T; 0 other;

alignment_scores:

Quality: 181.00 Length: 473

Ratio: 0.834 Gaps: 19

Percent Similarity: 45.877 Percent Identity: 20.930

alignment_block:

US-09-512-581-2 x AAV23545 ..

Align seg 1/1 to: AAV23545 from: 1 to: 6605

1036 SerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052

|||||

930 AGCCAGCGCAAGAAATTTCTCCCGCCAGCGCAATCTGGCCCAAGCAGTGGAGGA 979


```

566 LeuValSerPro.....ThrCysSer.....CysLys 574
|||:|||||
3935 CTAACCAACCACTACAGCTGGCTCTGCACAGCTAGCTGAATGGGAAATGCCAA 3984
574 sGlnAlaGlu.....GlyCysValArgGluIleThrLysLysLeuG 588
|||:|||||
3985 GACTTTGGAAAGAGTTGGCGCATGTGGCATGATGTTATGTCATACTTG. 4033
588 LysAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeu 604
|||:|||||
4034 .....GAGAAAGCAACACAGCTGGCTAAATGAGTAGAATTTAA 4072
605 LeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLe 621
|||:|||||
4073 CTTAAACCACTGAAACATCTCTGGCGAGCTGAGGAAATCTCTGAGGT 4122
621 uIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspG 638
|||:|||||
4123 GCTAGATTCACTTGAATAATTTGATGCGACATTCAGAGGATAACCCAAATC 4172
638 LuclyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLys 654
|||:|||||
4173 AG.....ATTCGC 4180
655 ValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheG 671
|||:|||||
4181 ATATTGGCACAGACC..... 4195
671 uSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAlaA 688
|||:|||||
4196 .....CTAACAGATGGCGGAGTCATGATGAGCTAATCAATGAG...G 4235
688 IaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPhePro 704
|||:|||||
4236 AACTTGAGACATTTAATCTCTGTTGGAGGAACACTACATGAAGAGCTGTA 4285
705 HisIleArgSerAlaLeuProValLeuHisHisLysSerLysLysLys 721
|||:|||||
4286 AGG...AGSCAAAGTTGCTTCAA..... 4306
721 yProProArgGlnAlaLysTyAlaIleHisCysIleHisAlaIlePheS 738
|||:|||||
4307 .....CAGAGCATCCAGT 4319
738 eSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSer 754
|||:|||||
4320 CTGCCAGGAGACTGNA.....AAATCC 4342
755 LeuAspProSerAsnLeuGluHisIleThrProLeuValThrIle... 770
|||:|||||
4343 TTA.....CACTTAATCCAGGAGTCCCTCACATTCAT 4374
771 ....GlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpL 786
|||:|||||
4375 TGACAGACAGTTGGCAGCTTATATTCAGACAGAGGTGGACGCGCTCAAA 4424
786 ySerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArg 802
|||:|||||
4425 TCCTTCAGGAAGCCAGAAATCCCAATCTGATTG..... 4459
803 LeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSe 819
|||:|||||
4460 .....ACAAGTCATGAGATCAG 4476
819 rProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL 836
|||:|||||
4477 TTGAGAAAGAAATGAAGAAA..... 4495
836 euLeuGlyMetLysAsnHisSerLysSerGlyThrSerThrLeuArg 852
|||:|||||
4495 ..... 4495
853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLys 869
|||:|||||
4496 .....CATAATCAGGGGAAGGAGCTGCCCAAGAGT 4527
869 sIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaI 886
|||:|||||
4528 CTGTCTCAGATTGATGTTGCACAGAAAAATACAAAGATGCTCCATGA 4577
886 leValLysLeuAlaGlnGluProCysTyHisGluIle..... 898
|||:|||||
4578 ACTTTCGATTATTCAGAAACACAGCCAAATTTGAGCTGCTACAGAA 4627
899 .....IleThrLeuGluGlnTyGlnLeuCysAlaLeuAlaIleAsnAs 913
|||:|||||
4628 ACTAAGATGATTTAGATGAAGTGAAGATGCACCTTGCCTGCATTTGAAAC 4677
913 pGluCysTyGlnValArgGlnValPheAlaGln.....L 925
|||:|||||
4678 AAGAGCTGGGAACAGGAAGTAGTACAGTCACAGCTAAATCATTTGTGTA 4727
925 ySLeuHisLysGlyLeuSerArgLeuArg..... 934
|||:|||||
4728 ACTTGTATAAAGTCTGAGTGAAGTGAAGTCTGAAGTGAATGTTGATA 4777
934 ..... 934
4778 AAGAGCTGGAGCTCAGATTGTACAGAAAAAGCAGACGGAATAATCCCAAGA 4827
935 .....LeuProLeuGluTyMetAlaIleCysA 944
|||:|||||
4828 ACTTGATGAAGAGTAACAGCTTTGAAATGTCATTATAATGAGCTGGGAG 4877
944 IaLeuCysAlaLysAspProValLysGluArgAlaHisAlaArgGln 960
|||:|||||
4878 CA.....AGGTAACAGAAAGAAACCAACAGTGGAGAAA 4912
961 CysLeuValLysAsnIleAsnValArgArgGluTyLeuLysGlnHisAl 977
|||:|||||
4913 TGCTTGAATTTGCCGTGAAGATGCGAAAGAA..... 4945
977 aAlaValSerGluLysLeuLeuSerLeuLeuProGluTyValValPro 994
|||:|||||
4946 .....ATGANTGCTTTGACAGANTGG..... 4966
994 yThrIleHisLeuLeuAlaHisAspProAspTyTrValLysValGlnAsp 1010
|||:|||||
4967 .....CTGGCAGCTACAGATATGGAATTGACAAAGAGATCAGCA 5005
1011 IleGluGlnLeuLysAsp.....ValLysGluCysLeuTrpPheValLe 1025
|||:|||||
5006 GTTGAAGGAATGCCGTAGTAATTTGGATTCTGAAGTTGCCCTGG..... 5047
1025 uGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgL 1042
|||:|||||
5048 .GGAAGGCTACTCAAAAAGAGATTGAGAAACAGAGGTGCACCTGAAGA 5096
1042 yMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAsp 1058
|||:|||||
5097 GTATCACAGAGGTAGAGAGCGCTTGAAAAACAGTTTGGGCAAGAGGAG 5146
1059 AlaLysMetAsnGluLysLeuTyThrVal..... 1068
|||:|||||
5147 ACCTTGGTGGAAAGATAAATCAGTCTTCTGAATAGTAAGTGAAGTGTGT 5196
1069 .....CysAspValAlaMetAsnIleIleMetSer...LysSerT 1081
|||:|||||
5197 CACTCCCGCAGCAGAGAGTGGTTAAATCTTTGTTGGTAATACCAGAAAC 5246
1081 hrThrTySerLeuGluSerProLysAspProValLeuProAlaArgPhe 1097
|||:|||||
5247 ACATGGAAACTTTTGACCAGAAATGTGACCACATC.....ACAAAGTGG 5290
1098 PheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyLeuProPr 1114

```


XX Muscular dystrophy gene - used for prepn. of probes, dystrophin
 PT polypeptide and antibodies for diagnosis and therapy of muscular
 PT dystrophy
 XX Disclosure; Fig 5; 69pp; English.
 XX The inventors claim an MD probe comprising a purified ss NA SQ which
 CC hybridises to at least a part of the MD gene; pure dystrophin (DS)
 CC polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
 CC probes are equal to or greater than 10b of one of 12 cDNA sequences
 CC deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd
 CC gene.
 XX
 SQ Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T; 7 other;

alignment_scores:

Quality: 181.50 Length: 1405
 Ratio: 0.305 Gaps: 57
 Percent Similarity: 42.420 Percent Identity: 17.722

alignment_block:

US-09-512-581-2 x AAN90338 ..

Align seg 1/1 to: AAN90338 from: 1 to: 12923

```

110 IleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAs 126
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2804 ATAAAGTCAGTTAAATAATTTGTAAGAT.....GAAGTCAA 2841

126 nArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnI 143
    |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2842 CCGCTATCAGTCTCAACCTCAATTTGACGATTTAAATTCAAAGCA 2891

143 leCysPheGlu.....LeuGluAspSerAsn 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2892 TAGCCCTGAAAGAGAGAGGACAGGCCATGTTCTCGGATGCGAGACTTT 2941

152 GluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnG1 168
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2942 GTGCGCTTTACAAATCATTTTAAAGCAAGTCTTTCTGATGTGCAGGCCAG 2991

168 yHisAsnGln.....LysValHisMetHisMetV 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2992 AGAAGAGAGCTACAGACAATTTTGACACTTTCCACCAATCGCCTATC 3041

178 alAspLeuMetSerSerIle..... 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3042 AGGAGACCATGAGTGCCTCAGGACATGGGTCCAGCAGTCCAGAAACCAA 3091

185 .....IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVa 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3092 CTCCTCCATACCTCAACTAGTGTACCGACTATGAAATCATGGGACAGAG 3141

199 lLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrA 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3142 ACTCGGGAATTCAGCTTTACAAATTTCTCTGCAAGACCAACAAAGTG 3191

216 spLeu.....AlaLysAlaLeuLeuLysArgThrAla 226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3192 GCCTATATCTCAGCACCACTGTGAAGAGATGTCCGAAGAAAGCGCC 3241

227 GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLe 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3242 TCTGAAATAGCCGGAATATCAATCAGAAATTTGAAGAAATTTGAGGACG 3291*

243 uGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu... 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3292 CTGGAAGAGAGCTCTCTCCAGCTGGTTGAGCATTTGTCAAAGCTAGAGG 3341

259 .....IleLeuGluLeuTyrAsnIleAspSerHisLeu.....LeuLeu 271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

3342 AGCAAAATGAATAAATCCGAAAATTCAGAAATTCAGATACATACAAACCTGAAG 3391
272 SerValLeuProGlnLeuGluPheLysSerAsnAspAsnGluG1 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3392 AAATGGATGCTGAAGTTGATGTTTTTCTGAAGGAG..... 3427

288 uArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAsps 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3428 .....GAATGCCTGCCTTGGGCAAT 3449

305 erGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArg 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3450 CAGAAATTCCTA.....AAAAGCAGCTGAACAGTGCAGACTT...TTA 3490

322 PheAsnAspIleHis...ValProIleArgLeuGluCysValLysPheAl 337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3491 GTCAGTGATATTCAGACAAATTCAGCCAGCTTAACAGAGTGTCAATGAAGG 3540

337 aserHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluT 354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3541 TGGCAGACAGATAAAGATGAAGCA..... 3565

354 yrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAspVal 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3566 .....GAGCCAGAGTTTGTCTCGAGA..... 3586

371 lIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAs 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3587 .....CTTGAGACAGAACTCAAGAACTTAACACTCAGTGGGA 3624

387 pHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgV 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3625 TCACATGTGCCACAGGTC.....TATG 3647

404 alArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3648 CCAGAAAGAGGCGCTTGAAGGA.....GGTTGGAGAACTGTAAAGC 3691

421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLy 437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3692 CTCAG.....AAAGATCTATCAGAGATGCACGAATGGATGAC 3729

437 sasLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuL 454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3730 ACAAGCT.....GAAGAAGAGT 3746

454 euValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThr 470
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3747 ATCTTGAGAGAGATTTT...GAATATAAACTCCAGATGAATTACAG... 3790

471 ThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAs 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3790 ..... 3790

487 nAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgH 504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3791 .....AAAGCAGTTGAAGAGATG..... 3808

504 isGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3809 .....AAGAGAGCTAAAGAAAGAGGCC 3829

521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPr 537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3830 CACAAAGAAAGAA.....GCGAAAGTGAACCTCTTACTGAGTCTGTAAA 3873

537 oAsp.....ProGlyLysAlaGlnAspPheMetLysLysP 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3874 TAGTGTCATAGCTCAAGCTCCACCTGTAGCACAA..... 3907

549 heThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGluVal 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3908 .....GAGGCCTTAAAAAAGGAAGTCTGAAACT 3934

```


1045GluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
 4112 CTATGATATATATGATGATATAAAAAA...AATATAGTACGATGAA 4158
 1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMe 1073
 4159 AATAATTATAAATTAAATGAAAGATG.....AACATATCTAT 4196
 1073 t..... 1073
 4197 GTCAATGATGAAGATATGATTCCTACGTTAAATAGTGAACATGGAATA 4246
 1074AsnIleIleMetSerLysSerThr..... 1081
 4247 ATTTTCCAGTTGTCACCGCAATTTATTAGAAAAAAGTACTTATATA 4296
 1082ThrTyrSerLeuGluSerProLysAspProValLeuProAl 1095
 4297 GATTGGAATTTATGATAGTAAATCTATGGCAGT..... 4332
 1095 aArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrL 1112
 4333TTTACAGAAGAAAAATATAATTTTGTAAATAATGAAATGATT 4375
 1112 eu.....ProProGluMetLysSerPhePheThrProGlyLysPro... 1125
 4376 TATTCATACTAAAAGATGGAGTTTAAATTTTCCAGGGTAAAAATCTG 4425
 1125 1125
 4426 TTTAATAATAAATTTTAAATGATATCAATCAGGATGGTGTTCCTTT 4475
 1126LysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSera 1141
 4476 TTTTAAAAATATGAATCTCTTTTAGGCACTTAATAA..... 4512
 1141 laGlyLysGlnSerGlnThrLysSerArgMetGluThrValSerAsn 1157
 4513TCAATATAGCTTAAACAGTAGAGAGTGTAAANAAT 4548
 1158 AlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLe 1174
 4549 AGTAATAATAATTTAGTAAATAAT.....AAGGTGATGA 4583
 1174 uAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSers 1191
 4584 TAATATTGGAATATGGAGATATCAATACAACAATGTTACAATTTGCGA 4633
 1191 erPro.....LeuProGlyLys.....LysSerAspLys 1200
 4634 GTGATGAACATATATCTACAAGGGAGATATACACGCAATCATTTTCT 4683
 1201 ArgAspAspSerAspLeuValArgSerGluLeuLysProArgGlyAr 1217
 4684 AGACAGCATATGATGTATCTCTTTTAAANAATTGAA...GCTAGAAGTAA 4730
 1217 gLysLysThrProValThr...GluGlnGluLysLeuGlyMetAspA 1233
 4731 AAAATATAGTGATATACTTATATAATAGGACAAAAAGTAATTTGGAGA 4780
 1233 spLeuThrLysLeuValGlnGlnLysProLysGlySerGlnArgSer 1249
 4781 ATGACATGAGACTTATTAATGAATGAA..... 4809
 1250 ArgLysArgGlyHisThrAlaSerGluSerAspGluGlnInTrpProGl 1266
 4810AATGTATGTAGTAACATAGATGTTAATGAATGGGAAGA 4847
 1266 uGlu.....LysArgLeuLysGluAspIleLeuGluAsnG 1278
 4848 TAAGGTAATGGTACATGTAATAGTGTGGTGATAAAGAGACTGAAAAAGA 4897

1278 luAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPro 1294
 4898 ATATGAAAGAAATAATCAAGAAATAATGAAGAAT..... 4935
 1295 LysProLeuGlyGlyThrProLysGluGluProThrMetLysThrSe 1311
 4936AATGAAAGAAATAATGAAGAATAA 4961
 1311 rLysLysGlySerLysLysSerGlyProProAlaProGluGluGluG 1328
 4962 TGAAGAAGAAATAATGAAGAATAAT.....GAAAAAGA 4993
 1328 luGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLys 1344
 4994 ATATGAAAGAAATAATCAAGAAATAATGAAGAAATAATCAAGAAAT 5043
 1345 GlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSe 1361
 5044 AATGAAGAA.....AATAATGAAGAAAA 5066
 1361 rSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgP 1378
 5067 TAATCATATAGAAAGAAATGAT..... 5088
 1378 roSerLysThrProSerProSerGlnProLysLysAsnVal 1391
 5089 ..ATAAGGATAATAATTCGGGCAAGTGAAGAAATAATATA 5127
 seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC98079
 seq_documentation_block:
 ID AAC98079 standard; cDNA; 1939 BP.
 XX
 AC AAC98079;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO:89.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antinefactive; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200055351-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05883.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587534/55.
 DR P-PSDB: AAB53322.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 disorders such as colon cancer -
 XX
 PS Claim 1; Page 532; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;


```

2766 T.....AAATCTTTAAAC.....ATGCTTCTCTGTTATTAATAGAAA 2803
499 InAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
2804 ATAAATATATCAAGAGAGAAATGAAAAAATCAAGTGTATATTAAT 2853
516 ProLysThrAspAlaSerValLysAlaIlePheSerLysVal..... 529
2854 AAAAAAAGAATGAGATATGTTTAATGTTGTTGAAGGTACATTGTCA 2903
530 .....MetValIleThrArgAsnLeuProAspProGlyLysAlaG 543
2904 CATGGCCAAATGAATAATCAAGATAATATATATACCAAGGG..... 2946
543 InAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIle 559
2947 .....AATATAAAAAAATAATGAAGAAGAATA 2973
560 ArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAl 576
2974 ACCAAACATGATGAA...TATATATCAAGGAGAGAAAAAATAATATAA 3020
576 aGluGlyCysValArgGluIle..... 583
3021 TAGTAAATGTATTAGAAATTTTGCATGACTATAAATATGAACAAGTTTGA 3070
584 .....ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPhe 597
3071 GTTACCATACGTTGGATGAAGACAAAAAATAATGATGATGAACAATTTA 3120
598 LeuGluMet.....IleLysPheLeuLeuGlu 606
3121 ATAGATATGAATAAATGAGCGGATTATTGAACCGTGAATGGTGTATTAA 3170
606 uArgIleAlaProValHisIleAspPheArgGluSerIleSerAlaLeuIle 623
3171 TATATTTATTGGATAGAAAGATAATACAGTAGGAAGGATATGGAGA 3220
623 ysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGly 639
3221 AAGAGATGAGAGAGGAGATGGAGAACAGATGGAGAGGAGATGGAGAAG 3270
640 ValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValle 656
3271 GTGATGGAGAAG.....GAGATGGAGAAGGATGAT 3299
656 uSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerL 673
3300 G.....GAGAAGGAGGTGGAGAGAAGAAT 3322
673 euLeuAlaCysLeu.....LysMetAspAspGluLys 683
3323 TGAATAAATGAGATGAACAATAGGATGAACAATAGGATGAACAATGAGATG 3372
684 ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
3373 AAAAAATGAA.....ATAAACATTTATAAAAAAATAATGAGATATATGTAGA 3416
700 uGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisL 717
3417 TAATGAT.....AAGAAGTAGAATTTGTAATGAGG 3448
717 ysSerLysLysGlyProProArgGlnAlaLysTyrrAlaIleHisCysIle 733
3449 AGAAGAAACTTATTTACCCATTTAATTAAGTCTGTGATACAT..... 3492
734 HisAlaIlePheSerSerLysLysGluThrGlnPheAlaGlnIlePheGluPr 750
3493 .....AAAAA 3497
750 oLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProL 767

```

```

3498 TATGAATATGTCAATTAATAATAAATTTGTATAA..... 3531
767 euValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAla 783
3532 .....GATGATTAT..... 3540
784 ProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAs 800
3541 .....AATAATATATTAAGAATAATATGTAGATAA 3569
800 nAspArgLeuProGlyLys..... 806
3570 TTCATGCTCTTCTCAAAAGAGAGAGAAATATATTCGACCTTTTATTCAAAT 3619
807 .....LysThrThrLysLeuTrpValProAspGluValSerProGlu 821
3620 TAAATAAGAAAGATAAAGTATGGAA..... 3645
822 ThrMetValLysIleGlnAlaIleLysMetValArgTrpLeuLeuGlu 838
3646 .....CGTTTTAA 3653
838 yMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuT 855
3654 TATAAAGAATAAT.....ATTA 3670
855 hrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSer 871
3671 AGACAATAATACATAATGAA..... 3690
872 LysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVally 888
3691 .....GAGATGAAAAAGATAATATCAAACTATTAATAAAAAATGTTTTT... 3732
888 sLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrG 905
3733 .....TTTAATCGATATG 3757
905 InLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnVal 921
3758 AA.....AATTTTTTAATAATCATTTTAACATATAATTTTCCAAAAAT 3801
922 PheAlaGlnLysLeuHisGlyLysGlyLeuSerArgLeuArgLeuProLeuGlu 938
3802 GATTTATTTAAATTTATCATCAAAAGTAAGTATGAATAATAAGGAATTT 3851
938 uTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgA 955
3852 GTATATTGCT..... 3861
955 rgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
3862 .....AATAAACATATAATAATAATATATATATGAT 3888
972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuPr 988
3889 TATATCAATAAA.....TTATATAATCAAAATATA..... 3918
988 ogluTyrValValProTyrThrIleHisLeuLeuAlaHisAspPro.... 1003
3919 .....TATACATTTAAATATCAGGTAGCTTAATATAGATAATGATCATCATA 3964
1004 .....AspTyrValLysValGlnAspIleGlu 1012
3965 TATGTAAGAAAGGGGAGGTTGGATTATATAAATATGAATATATCAAAA 4014
1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMe 1029
4015 GAATGTAATAATAGGAAGAC...AAAACATATTTAAATAAAATATTTC 4061
1029 talLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal.... 1044
4062 TTATAGAGAAAAAGGATGCACGCTTTTATTAATGCGAAATTTGGTT 4111

```



```

Quality: 193.00          Length: 1597
Ratio: 0.256            Gaps: 78
Percent Similarity: 44.709 Percent Identity: 17.095

alignment_block:
US-09-512-581-2 x AAA70106 ..

Align seg 1/1 to: AAA70106 from: 1 to: 7458

2 AlaHisSerLysThrArg.....ThrAsnAspGlyLysI1 13
   ::::|||||:::
1336 AGTTATTCAAAAGTGAAAGACTATTATATACAATAATAAAACAGTAA 1385
   ::::|||||:::

13 eThrTyProPro.....GlyValLysGluLeuSerAspLysIleSerL 28
   ::::|||||:::
1386 TACGTTTCATTCTCTATATTTTCTTAAATAAGTAGGAGACAAGTTTAA 1435
   ::::|||||:::

28 ysGluGlu.....MetValArgArgLeuLysMetValVal... 39
   ::::|||||:::
1436 ACTCTGGAGATATATATGATGTATATAATAAAATAAAATGCTTATA 1485
   ::::|||||:::

40 .....LysThrPheMetAspMetAspGlnAspSerGluGluGluL 53
   ::::|||||:::
1486 CATGATAACAAGATATATCTAATATGATTCTTATAAATTAACAGAA 1535
   ::::|||||:::

53 sGluLeuTy..... 56
   ::::|||||:::
1536 ACATTATTATAGCACTTCTAATATATAAATTTATTATATAATAATAG 1585
   ::::|||||:::

57 .....LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHis 71
   ::::|||||:::
1586 AGGTATAGATAATGGTCTACATTTTCTCAATATCAATAAATATGAT 1626
   ::::|||||:::

72 ProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIle 88
   ::::|||||:::
1627 .....TGTCTGTTAAATCA..... 1641
   ::::|||||:::

88 eArgIleTyAlaProGluAlaProTyThrSerProAspLysLeu... 104
   ::::|||||:::
1642 .....AATCCACCATATAAGAGTATCTCGTTAAATTAATA 1675
   ::::|||||:::

104 ysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAsp... 119
   ::::|||||:::
1676 ACAATGTTTTTTTATATAAAGACGAAAGAAATAGTAGTAATAATAAT 1725
   ::::|||||:::

120 .....ThrLysSerProGlnPheAsnArgTyThrPheTy 130
   ::::|||||:::
1726 AATAATAATAATATTAGTAGTAGTAGTAGTAGTAGTAAAAAATAATCA 1775
   ::::|||||:::

130 rLeuLeuGluAsnIleAlaTrpValLysSerTyArgIleCysPhe... 146
   ::::|||||:::
1776 CGTGATCATTAAT...AAAAAATATCATCTTATATATATTCATTATAAG 1822
   ::::|||||:::

146 lLeuGluAspSerAsnGlu..... 152
   ::::|||||:::
1823 AAAGAAGATTTCGTTTAAAGAGATTTTTTATTTCACAGAGAAATTT 1872
   ::::|||||:::

153 .....llePheThrGlnLeuTyArgThrLe 161
   ::::|||||:::
1873 TPACCATCAAAAAAGATACTTGTTGTTTTTAATGAAAGACAAAGATCT 1922
   ::::|||||:::

161 uPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMet 178
   ::::|||||:::
1923 TTTT.....GAAAAAGCTAATGAACAT...A 1945
   ::::|||||:::

178 aIAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGlu 194
   ::::|||||:::
1946 TTAATGCGTTCTCTCTTTTAAATAACATCATGATATTTCT..... 1989
   ::::|||||:::

195 LeuLeuAspThrValValLeuValAsnLeuValProAlaHisLysAsnLeu 211
   ::::|||||:::
1990 .....TCACATTCAAGCTGATAA 2006
   ::::|||||:::

```

[illegible]


```
1079 .....LysSerThrThrTyrSerLeuLeuSerProLysAsp 1090
      ::::: ::::: ::::: :::::
9429 ACCGAAACACATCGAAACCTTTGATCAGAACATAGAACAAATCACAAG 9380
      ::::: ::::: ::::: :::::
1091 ProValLeuProAla....ArgPhePheThrGlnProAspLysAsnPhse 1106
      ::::: ::::: ::::: :::::
9379 TGGATCATTCATGCAGATGCAACTTTTAGATGAGTCTGAAAGAAAGAAACC 9330
      ::::: ::::: ::::: :::::
1106 rAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThrProG 1123
      ::::: ::::: ::::: :::::
9329 ACAACAAAGGAAGACATCTCTAAACGCTTTAAAGGCTGAATGAATGACA 9280
      ::::: ::::: ::::: :::::
1123 LyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSer 1139
      ::::: ::::: ::::: :::::
9279 TGGCGCCAAAGGTGGACTCCACACGTCACCAAGCAGCAAAATTTGATGCA 9230
      ::::: ::::: ::::: :::::
1140 SerAlaGly.....LysGlnSerGlnThrly 1148
      ::::: ::::: ::::: :::::
9229 AACCGCGGTGACCACTGCAGGAAAGTAGAGCCCAAAATCTCTGAGCT 9180
      ::::: ::::: ::::: :::::
1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProS 1165
      ::::: ::::: ::::: :::::
9179 CAACCGTCGATTGCGAGCTATTCTCACAGAAATTAAGACTGGAAGGCT 9130
      ::::: ::::: ::::: :::::
1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHis 1181
      ::::: ::::: ::::: :::::
9129 CCATTCTCT.....TTGAAGGAATTGGAGCAG 9104
      ::::: ::::: ::::: :::::
1182 SerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLys..... 1196
      ::::: ::::: ::::: :::::
9103 TTAACTCAGATATACAAAATTTGCTTGAACCACTGGAGGCTGAAATTC 9054
      ::::: ::::: ::::: :::::
1197 .....LysSerAspLysArgAspA 1203
      ::::: ::::: ::::: :::::
9053 CGAGGGGTGAATCTGAAGAGCAAGACACTTCAATAAGATATGAGTGAAG 9004
      ::::: ::::: ::::: :::::
1203 spSerAspLeuValArgSerGluLeuGluLeuProArgGlyArgLysLys 1219
      ::::: ::::: ::::: :::::
9003 ACAATGAGGTACTCTAAATGAATGTTGCAAGAGGAGGACAACTTACAA 8954
      ::::: ::::: ::::: :::::
1220 ThrProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrly 1236
      ::::: ::::: ::::: :::::
8953 CAAGAATCAGATGAGAGAGGAGGAGGAATAAAGATAAAACAGCA 8904
      ::::: ::::: ::::: :::::
1236 sLeuValGlnGluGln.....LysProLysGlySerGlnArgS 1249
      ::::: ::::: ::::: :::::
8903 GCTGTTACAGACAAACATAATGCTCTCAAGGATTTGAGGTCTCAAGAA 8854
      ::::: ::::: ::::: :::::
1249 erArgLysArgGlyHsThrAlaSerGluSerAspGluGlnIntrpPro 1265
      ::::: ::::: ::::: :::::
8853 GAAAAAAG.....GCCCTAGAAATTTCTCACCAGTGGTAT 8819
      ::::: ::::: ::::: :::::
1266 GluGluLysArgLeuLysLysAspIleLeuGluAsnGluAspGluGlnAs 1282
      ::::: ::::: ::::: :::::
8818 CAGTACAGAGGCGGCTGATGATCTCTGAAATCTGCTGATGAATTT... 8771
      ::::: ::::: ::::: :::::
1282 nSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyG 1299
      ::::: ::::: ::::: :::::
8771 ..... 8771
1299 LyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySer 1315
      ::::: ::::: ::::: :::::
8770 .....GAA 8768
      ::::: ::::: ::::: :::::
1316 LysLysLysSerGlyProProAlaProGluGluGlu.....GluG 1329
      ::::: ::::: ::::: :::::
8767 AAAAATTAGCCAGGCTACTGACCCAGGAGATGAAGAAATTAAGGA 8718
      ::::: ::::: ::::: :::::
1329 uGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnH 1346
      ::::: ::::: ::::: :::::
8717 AATTGATCGTGAATG.....CAGAAAGAAAGAGGAGGAGCTGA 8680
```

```
1346 IsArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSer 1362
      ::::: ::::: ::::: :::::
8679 ATGCAGTCGCGCAGGCAAGCTGAGGCTTGTCTGAGATGGGCGCCGAATG 8630
      ::::: ::::: ::::: :::::
1363 AlalleGluSerThrGln 1368
      ::::: ::::: ::::: :::::
8629 CGAGTGGAGCCCAACTCAG 8612
      ::::: ::::: ::::: :::::
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA70106
seq_documentation_block:
ID_ AAA70106 standard; DNA; 7458 BP.
XX
AC AAA70106;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:239.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN W020025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 462-464; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 7458 BP; 3523 A; 514 C; 1174 G; 2247 T; 0 other;
alignment_scores:
```


11150 CAGTTTAAATAGT.....GTTAATGAAGGTGGGCAGAAGATAA 11113
602 yspHeLeuLeuGluAArgIleAlaProValHisIleAspThrGluSerIle 618
11112 AGAGTGAAGCTGAACGTGAGTTGGATCCAGACTGCGAGAGAGAA..... 11069
619 SerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAs 635
11068CTTAGAGAGCTTACACTCACTGAGTGGATCATATGC..... 11033
635 pGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGlu. 651
11032CGCCAGGTCTACACCAGAAAGGAGCTTAAAGCGAGGTTGGATA 10987
652LeuLeuLysValLeuSerPheThrHisProIleSerPhe 664
10986 AACCGTAAAGCCCTCCAAAAGAGTCTATCAGAGATGCATGAGTGGATGACA 10937
665 HisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAs 681
10936 CAGCTGAAGAAGATATCTAGAGAGATTTTGAATATAAACTCCAGA 10887
681 pGlu.....LysValAlaGluAlaAlaL 689
10886 TGAATTACAGACTGCTGTTGAAGAAATGAAGAGAGCTAAAGAAAGAGGCAC 10837
689 euGln.....IlePheLysAsnThrGlySerLys 698
10836 TACAAAAGAAACTAAAGTGAACCTCTTACTGAGACTGTAAATAGTGA 10787
699 IleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHi 715
10786 ATAGCTCAGCTCCACCTCAGCAGCAAGAGGCTTAAAAAGGAAGTTGA 10737
715 sHisLysSerLys.....LysGlyProP 723
10736 AACTCTGACCACCACTACCAATGGCTGTGCACCAGGCTGAATGGAAT 10687
723 roArgGlnAlaLysTyraIleHisCysIleHisAlaIlePheSerSer 739
10686 GCAAACTTTGGAAGAAGTTTGGGCATGTGGCATGAGTTATTGTCTAT 10637
740 LysGlu..... 741
10636 TTAGAGAAACCAACAAGTGGCTCAATGAAGTAGAATTGAAACTTAAAC 10587
742ThrGlnPheAlaGlnIlePheGluP 750
10586 CATGGAAATGTTCTCTGAGGACCTGAGGAATCACTGAAGTGTAGAT 10537
750 roLeu.....HisLysSerLeuAspProSerAsnLeuGlu 761
10536 CTCTGAAATCTGATGCATCATTCAGTACAGAGAGAACCAATCAGATTCTG 10487
762 HisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPr 778
10486 CTATTGGCACAGACTTACAGATGAGGAGTCTATGATGAAGTATGATCAA 10437
778 oAspGln.....PheAlaAlaProThrLysSerTrpValAlaThrP 792
10436 TGAGGAGCTTGACAGCTTTAATCTCGTTGGAGGGAACATCATGAAGAGG 10387
792 heIleValLysAspLeuLeuMetAsnArgLeuProGlyLysLysThr 808
10386 CTGTGAGGAAACAAAGTTGCTTGAACAGAGTATCCAGTCTGCCAGGAA 10337
809 ThrLys.....LeuTrpValProAspG 816
10336 ATTGAAAGTCCTTGCACTTAATTCAGGAGTCCCTTGAATTCATTGACAA 10287
816 uGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetv 833
.....

10286 GCAGTTGGCAGCTTATATCACTGACAAGGTGGATGCAGCTCAAATGCCTC 10237
833 alArgTrpLeuLeuGlyMetLysAsnHisSerLysSerGlyThrSer 849
10236 AGGAA.....GCCAGAAAAATCCAATCAGATTTTGACAAGTCAT 10199
850 ThrLeuArgLeuLeuThrThrIleLeuHisSerProAspMetSerArgLeuAlaAlaG 866
10198 GAGATAGTTTGAAGAAATCAAGAACATACACGGGGAGGATGCCAA 10149
866 uGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaG 883
10148 CCAAGGGTCTTTTCAAAATTCATGTTGCACAGAAAAAATTTACAAGATG 10099
883 lySerAlaIleValLysLeuAlaGlnGluProCysTyThrHisGlu..... 897
10098 TCTCCATGAATTTTCGATTTATCCAAAACACGCCAATTTTGAACAACGT 10049
898IleIleThrLeuGluGlnTyThrGlnLeuCysAl 908
10048 CTAGAGAAAGTAAAGATGATTTTAGATGAAGTCAAGATGCTTTTGCCTGC 9999
908 aLeu.....AlaIleAsnAspGluCysTyThrGlnValArgGlnValP 922
9998 ATTGGAACCAACAGAGTGTGAACAGGAAGTAATTCAGTCACAACCTAAGTC 9949
922 heAlaGlnLysLeuHisLysGlyLeuSerArgLeuArg..... 934
9948 ATTGTGTGAACCTGTATATAAAGCCTGAGTGAAGTCAAGCTGAAGTGGAA 9899
934 934
9898 ATGCTGATTAACCCGAGCGCTCAAAATGTACAGAAAAGCAGACAGAAA 9849
935LeuProLeuGluTyThrMeta 941
9848 TCCCAAAGAGCTTGTCAACGAGTAAACAGCTTTGAAATTCATTACAATG 9799
941 laIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
9798 AGTTGGGTGCG.....AAGGTAACAGAGAGAAAGCAACAG 9764
958 AlaArgGlnCysLeuValLysAsnIleAsnValArgGlnTyThrLeuLy 974
9763 TTGGAATGCTTGAAGTTGTCCTCGTAGATGAGAAAGGAA..... 9722
974 sGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrv 991
9721ATGAATGCTTTAACAGATGG. 9701
991 alValProTyThrIleHisLeuLeuAlaHisAspProAspTyThrValLys 1007
9700CTGGCAGCACACATACAGAATTTGACGAAG 9671
1008 ValGlnAspIleGluGlnLeuLysAsp.....ValLysGluCysLeuTr 1022
9670 AGATCAGCTGTGAAGAAATGCCAAGTAATTTGGATTCCTGAAGTTGCCTG 9621
1022 pPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaP 1039
9620 G.....GGAAAGGCTACTCAAAAAGAGATTGAGAAACAGAGGCTC 9580
1039 heIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGly 1055
9579 ACTTGAAGAGTGTACAGAAATAGGAGAGTCTTTGAAATGGTGTGGGC 9530
1056 ProAspAlaLysMetAsnGluLysLeuTyThrVal..... 1068
9529 ARGAAAGAACCTTGGTAGAGATAAAGTCTTCTGAACTGAACTG 9480
1069CysAspValAlaMetAsnIleMetSer. 1078
9479 GATAGCTGTACCTCCAGAGTAGAAGAAATGGCTAAATCTTTTGTGGAAT 9430


```

2842 GGCTTCTCATCATGTTGAAATGGAACTCAATTC.....A 2876
890 laclnGluProCysTyrHisGluIleThrLeuGluGlnTyrGlnLeu 906
2877 TCGAGGAACAACATTACAGGCCGACAAATCTTTAAACAAATGAAAA 2926
907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAl 923
2927 GAAGACCAACAACCTTCAGGGAACATAAATGAGCTAAACCAATCAAGTCA 2976
923 aGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrM 940
2977 ACAGAAG...AAAAACAATAATTGAAGCATCCCAAGGAGAGCTTAAATTC 3023
940 etAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
3024 CAGTTTACAGAGACAGAGCTTGAGAAATAAATACAGCAGCAGTTAAACA 3073
957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgGluTyrLe 973
3074 CAGGCAGCCGGAACCTT..... 3091
973 uLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuProGluT 990
3092 .....GCAGCAGAGAAGAAATAATATCAGTATTACAAAACAACCT 3131
990 yrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVal 1006
3132 ATGAAAAAGTCAGGAAACTTCAACACAGCTTCAATCTGATTTATGGG 3181
1007 LysValGlnAspIle.....GluGlnLeuLysAspValLysG 1019
3182 AGGGAATCTGAACCTTACCCACCAGCAGGATCTTAAGTCTGTAGAACA 3231
1019 uCysLeuTyrPheValLeuLeuLeuMetAlaLys..... 1031
3232 GAAGCTTTCTAGCAGCAGGAGGACTTGATTTCAACAGAAATCAAATTC 3281
1032 ..AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1047
3282 GAATCAAAAT.....AAATGATTCAAGAACTG 3310
1048 LysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGlu 1064
3311 AAGACTGCCAAGCTACATTTGGAGCAGGATTCAGCAAGAAGAAACAGCA 3360
1064 sLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerT 1081
3361 ATTGCAGGAGCGATGTAAAGCAGCTACAAGACATTCAGAAGAAAGTCAC 3410
1081 hrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhe 1097
3411 TG..... 3412
1098 PheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1114
3413 .....AAGAAAAAGAACTGGTAATGAGAGCTCTAAATTC...GC 3450
1114 oGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnVal 1131
3451 AGAGATAGAAGAAAT.....AAATGTACAGAAAGAAAGAAA 3488
1131 euGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr 1147
3489 TCATAAATAACGAAAGAACTC..... 3511
1148 LysSerSerArgMetGluThrValSerAsnAlaSerSerSerAsnPr 1164
3512 AATCCCAACAACTAGAACGATA..... 3535
1164 oSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
3535 ..... 3535

```

```

1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLys 1197
3535 ..... 3535
1198 SerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysPr 1214
3536 .....AAGGAGATAACAAATCT 3552
1214 oArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyM 1231
3553 TAAAGATGCTAAACAGACTTCTAATTCAGCAGAAATTTAGAACTTCAAGGAA 3602
1231 etAspAspLeuThrLysLeuValGlnGlnLysProLysGlySerGln 1247
3603 AGCGGACTCCCTGAAGCAGCTGTGAACAGAGAGAGAAATCAGCAG 3652
1248 ArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
3653 ATACTAAAA.....GACCAGGTGAA 3672
1264 proGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluG 1281
3673 AAGGAAGAAGAGGAGCTGAAGAAGAATTTATTGAGAAGAAGCTTAAGT 3722
1281 lnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeu 1297
3723 TGCATTC..... 3730
1298 GlyGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysG 1314
3731 .....GAAATAAAAGAAAGAACTAGG 3753
1314 ySerLysLysLysSerGlyProProAlaProGluGluGluGluGluG 1331
3754 AATGAAGAAG.....CATGAAGAAATGAGG 3779
1331 luArgGlnSer.....GlyAsnThr 1337
3780 CTAAATTCACCATGCAGATTACAGATTAAATGAAACTTAGGCCTGTG 3829
1338 GluGlnLysSerLysSerLysGlnHisArgValSerArgAlaGlnG 1354
3830 AAGAAGGAGTGGCAATCTAGTCAACGAGAGTTAGTGAGTTGAGAAACA 3879
1354 nArgAlaGlu...SerProGluSerSerAlaIleGluSerThrGlnSerT 1370
3880 AACGGATGACTTACGGGGTGAAATTCAGTATTAGAACCAACGGTTTCA 3929
1370 hrProGln.....LysGlyArg 1375
3930 ATAACTCAAGATGAAGGAGAGCAGCTACTGAAAGATGCTTTAAAGGAGAA 3979
1376 GlyArgProSerLys 1380
3980 GTGGAATAGAAAAAG 3994

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:AA27558

seq_documentation_block:

ID AA27558 standard; cDNA; 19307 BP.

XX AA27558;

XX 07-AUG-1996 (first entry)

XX Shuttle vector pAdDel.CMVdys.

XX Adenovirus type 5; Ad5; vector; gene therapy; gene transfer; helper virus; dystrophin; muscular dystrophy; ds; cyclic.

XX OS Synthetic.

XX

773 leAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpVal 789
3080AGGACGTGGATC 3091
790 AlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLy 806
3092 CAG.....CA 3096
806 slyThrThrLysLysLeuTrpValPro.....AspGluGluV 818
3097 GTCAGAAAGCAAACTCTGTCTACCTTATCTTAGTTACTGAATGAAA 3146
818 alSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArg 834
3147 TAATGAGGACACACTCGGGAATATACAGCTCTGCAAAAGTCTTTGAAA 3196
835TrpLeuLeuGlyMetLysAsnHisSe 844
3197 GAGCAACAAATGGCTTCAACTATCTGAGTGACACTGTGAAGGAGATGC 3246
844 rLysSerGlyThrSerThrLeu.....ArgLeuLeuThrIleLeuH 859
3247 CAAGAAGCACCTTCAGAAATATGCCAGAAATATCTGTCAGAAATTTGAAG 3296
859 isSerAspGly.....AspLeuThrGluGlnGly 868
3297 AGATTGAGGGGCTCGAGAACTTCTCCCGAGTTGTGGAAAGCTGC 3346
869 LysIleSerLysProAspMetSerArgLeuArgLeu..... 880
3347 CAAAAGCTAGAACACATATGAATAAATTCGAAATTTTCAGAAATCACAT 3396
880 880
3397 AAAAACCTTACAGAAATGGATGCTGAAGTTGATTTTCTCTGAAAGAG 3446
881AlaAlaGlySerAla...IleValLysLeuAlaGlnGluPro 893
3447 AATGGCTCGCCCTGGGGGATCGTGAATCTCTGAAAAACAGCTCAACAA 3496
894 CysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAl 910
3497 TGCAGATTTTGTGTGATATTCAAACAATTCAGCCAGTTTAAATAG 3546
910 alIleAsnAspGluCysTyrGlnValArgGln.....ValPheA 923
3547 TCTTAATGAAGGTGGCAGAGAATAAGAGTGAAGCTGAACCTTGAGTTTG 3596
923 laGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyr 939
3597 CATCCAGCTGGAGACAGAACTTAGAGAGCTTAACACTAGTGGGATCAC 3646
940 MetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAl 956
3647ATATGC..... 3652
956 aHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrL 973
3653CGCCAGGTCTACACCAGAAAGAGCCCT 3680
973 euLys.....GlnHisAlaAlaValSerGluLysLeuLeuSerLeu 986
3681 TAAAGCAGGTTGGATTAACCGTAAAGCTCCCAAAAGATCTATCAGAG 3730
987 LeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspPr 1003
3731 ATGCATGAGTGGATGACACAACTGAAGAAGATAATCTCAGAGAGATTT 3780
1003 oAspTyrValLysValGlnAsp.....IleGluGlnLeuL 1015
3781 TGAATAT...AAAACCTCCAGATGAATTACAGACTGCTGTTGAAAGAAATGA 3827

1015 ySaspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLys 1031
3828 ACAGACCTAAAGAA.....GAGGCACCTACAAAAGAA 3859
1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLy 1048
3860 ACTAAA.....GTCAAACCTCTTACTGAGACTGATAA 3891
1048 sGlnThrLysAspAlaGlnGlyProAspAspAlaLys.....MetAsnG 1063
3892 TAGTGTA...ATAGCTTCAGCTCCACCTCCAGACAAAGAGCGCTTAAAAA 3938
1063 LuLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLys 1079
3939 AGGAACCTGAAACTCTG..... 3955
1080 SerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaAr 1096
3955 3955
1096 gPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuP 1113
3956ACCACCACTAC.... 3967
1113 roProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsn 1129
3968CAATGGCTGTCCACCAGGCTGAATGGAAATGCAAACTTTGGAA 4012
1130 ValLeuGlyAla.....ValAsnLysProLeuSerSerAlaG 1142
4013 GAAGTTGGGCATGTTGGCATGAGTTATTGTCATATTAGAGAAAGCAAA 4062
1142 yLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlas 1159
4063 CAATGGCTCAATGAAGTAGAATGAACTTAAACCATGGAAATGTT. 4111
1159 erSerSerSerAsnProSerProGlyArgIleLysGlyArgLeuAsp 1175
4112CCTGCAGGACCTGAGGAATCACTGAAGTGTAGAA 4147
1176 SerSerGlu.....MetAspHisSerGluAsnGluAspTyrThrMetSe 1190
4148 TCTCTTGAATAATCTGATGCATCATTCAGAG...GAGAACCCAAATCAGAT 4194
1190 rSerProLeuProGlyLysLysSerAspLysArgAspSerAspLeuV 1207
4195 TCGTCTATTGGCACAGACTCTTACAGATGAGAGTCATGGATGAACTGA 4244
1207 alArgSerGluLeuGlu.....LysProArgGlyArgLysLysThrPro 1221
4245 TCATGAGGAGCTTGAGACGTTTAAATCTCTGTTGGAGGGAACCTACATGAA 4294
1222 ValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysLeuVa 1238
4295 GAGGCTGTGAGGAACAAAGTTG..... 4318
1238 lGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHist 1255
4319 .CTTGAACAGAGATATCCAGTCTCCCGAGGAATGAAAGTCCTTCGACT 4367
1255 hrAlaSerGluSer..... 1259
4368 TAATTCAGGAGTCGCTTGAATTCATTACAAAGCAGTTGGCAGCTTATATC 4417
1260AspGluGlnGlnTrpProGluGlu...LysArgLeuLy 1271
4418 ACTGACAAGGTGGATGCAGCTCAATGCCTCAGGAAGCCAGAAATCCA 4467
1271 sGluAspIleLeuGluAsnGlu.....AspGluGlnAsnSerProp 1285
4468 ATCAGATTTTGCAGCAAGTCATGATAAGTTAGTAAAGAAATG. 4507
1285 roLysLysGlyLysArgGlyArg 1292

1366 TGCTCATGAGGATTCATGATGATCTGACATCTCATCAAGGACTTGTG 1415
292ValValLysLeuLeuAlaLysMetPheGly.....Ala 302
|||:||||| |||:|||||
1416 GTAATGTTCTACGTTAGGAAGTCAACTAGTTAGTGGAAAGGGAATATCA 1465
|||:||||| |||:|||||
303 LysAspSerGluLeuAlaSerGlnAsnLys.....ProLe 314
|||:||||| |||:|||||
1466 GAAGATGAAGAAGCTCAAGTGAAGAACAATAATGAATCTCTAAATTCAG 1515
314 uTrpGlnCysTyr.....LeuGlyArgPheAsnA 324
|||:||||| |||:|||||
1516 ATGGGAATGCTCAGGGTAGCTAGCATGGGAAACAAAGCAAAATACACA 1565
324 spIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
|||:||||| |||:|||||
1566 AAGTCTTAATGATCTCCAGATCCAGAAATTAAGAAGAACTAGATGACTGG 1615
341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu..... 353
|||:||||| |||:|||||
1616 TTAACAAAACCTGAAGAGAGAACTAAGAAATGGAGGAAGCCCTTTGG 1665
354TyrLeuLysValArgSerHisAsp..... 361
1666 ACCTGATCTTGAAGATCTAAAAATGCCAAGTACAACAACATAACGCTTTC 1715
362ProGluGluAlaIleArg.....Hisasp 369
1716 AGAAGATCTAGAACAGGAGCAGGTCAGGTCACCTCGCTCACTCACATG 1765
370 ValIleVal.....SerIleValThrAlaAlaLys 379
|||:||||| |||:|||||
1766 GTAGTAGTGTGTTGATGAATCCAGCGGTGATCATGCAACAGCTCTTTGGA 1815
379 slyAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluA 396
1816 AGAACAACTTAAGGTACTGGGAGATCGAGGGGCAAAATATCTGC.....A 1859
396 rgThrLeuAspLysArgTTPArgValArgLysGluAlaMetMetGlyLeu 412
|||:||||| |||:|||||
1860 GATGGACTGAACACCGCTGGATGTTTACAGATATCTTCTTAAATGG 1909
413 AlaGlnIleTyrLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAl 429
1910 CAGCATTTTACTGAAGACAGTGCCTTTTACT..... 1942
429 aAlaLysGlnIleAlaTTPileLysAspLys.....LeuLeuHisI 443
1943ACATGGCTTTCAGAAAAGAGATGCAATGAAGAACA 1979
443 leTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePhe 459
|||:||||| |||:|||||
1980 TTCAGACAAGTGGCTTAAAGATCAAAATGAATGATGTCAGTCTTCAC 2029
460 AlaGlnTyrMetValProHisAsnLeuGlu.....ThrThrG 472
2030 AAAATATCTACTTTAAAAATAGATAGTAAAGAAAAGAAAGCCACCATGA 2079
472 uArgMetLysCysLeuTyrTyr.....LeuTyrAlaThrLeuAspLeuAsnA 488
2080 AAAACTAAGTTCACATCAAGATCTACTTTCCGGCAGCTGAAAAATAAGT 2129
488 laValLysAlaLeuAsnGluMetTTP.....LysCys 498
2130 CAGTGACTCAAAAGATGGAATCTGGATGGAACAACTTTCACACAGCTGG 2179
499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLys... 514
2180 GACAATTTAACCCCAAAAAAGAAAGAGTTTCAGACAAATTTTCACAGGC 2229
515GlnProLys.....ThrAspAlaSerValLysAlaIleP 526
|||||:||||| |||:|||||

2230 TGTCACCACTCAACCATCCTTAACACAGACAACTGTA.....ATGG 2273
526 heSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAla 542
|||:||||| |||:|||||
2274 AAACGGTAACTATGTTGACCAACAAGGAAACAATCATGTGTAACACATGCC 2323
543 GlnAspPheMet.....LysLysPheThrGlnValLeuG 554
|||:||||| |||:|||||
2324 CAAGACGAACCTTCCACCACCTCTCTCAAAAAGAGAGGACAGATAACTGT 2373
554 uAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThrC 571
2374 GGATCTGAA...CTCAGGAAAAGGTTGGATGTCGATATAACTGAACTTC 2420
571 ysSerCysLys..... 574
2421 ACAGTTGCATTACTCTCTCAGAAGCTGTATTACAGAGTTCTGAATTTCCA 2470
575GlnAlaGluCysValArgGluIleThrLysLysLeuGlyAs 589
|||:||||| |||:|||||
2471 GTCTAACGAAAAGAGCAACATCTCAGACTTGCAGAAAAAAGTCAATGC 2520
589 n.....ProLysGlnProThrAsnProPheLeuGluMet..... 600
2521 CATAGCAGCAAGAAAAGCAGAGAAAGTTCAAGAACTGCAAGATGCCAGCA 2570
601IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
2571 GATCAGCTCAGGCCCTGGTGGAAACACATGCAAAATGAGGTTGTTAATGCT 2620
616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyTh 632
2621 GAAAGTATCAGACAAGCTTCAGAACAACTGAAC..... 2653
632 rAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaG 649
2654ACGCGGT 2660
649 lYLeuGluLeuLysValLeuSer..... 657
2661 GCACAGAATCTGCCAATTTGCTGAGTGAGAGATTAACTGGCTAGAGTAT 2710
658 PheThrHisProIleSerPheHisSer..... 666
2711 CAACACCAACATCATCTACCTTTTATAATCAGCTACACAATTTGGAACAGAT 2760
667 AlaGluThrPheGluSerLeuLeu.....AlaCysLeuLysMetAspA 681
2761 GACACTACTGCCGAAACTTGTGTAACCCAGCTACCCACCTATCAG 2810
681 spGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySer 697
2811 AGCCAAACAGCAATTAACCCAGTTAAAAATTTGTAAGGATGAAGTCAAC 2860
698 LysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValle 714
2861 AGATTGTACAGCTCTTACGCTCAAAATTTGACAAATTAATAATTCAGAGTCT 2910
714 uHisHisLysSerLys...LysGlyPro.....ProArgGlnA 726
2911 ACACACTGAAGAAAGGAGCAGGGCCCAATGTTCTGGATGCAGACTTTG 2960
726 laLysTyrAlaIleHisCysIleHisAlaIlePheSer.....Ser 739
2961 TGCCCTTTACTAATCATTTTAACCAAC...ATCTTGATGTTGTGAGGGCC 3007
740 LysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAs 756
3008 AAAGAGAAAGAGCTACAGACAATTTTGAC.....ACTTTACC 3045
756 pProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773
3046 ACCAATGGCCTATCAGGAGCAATGAGTAGCATC..... 3079

4169 AGATCCTTTCGAGGATTATTCCTCTCACCAAAATCGAGACCTCGAGAG |||||
4170
4171
4172
4173
4174
4175
4176
4177
4178
4179
4180
4181
4182
4183
4184
4185
4186
4187
4188
4189
4190
4191
4192
4193
4194
4195
4196
4197
4198
4199
4200
4201
4202
4203
4204
4205
4206
4207
4208
4209
4210
4211
4212
4213
4214
4215
4216
4217
4218
4219 TCCGATATTGTATTATTTAGCTACCAGAAGGAGAGGTTTCGCTTTCACGGATA
4220
4221
4222
4223
4224
4225
4226
4227
4228
4229
4230
4231
4232
4233
4234
4235
4236
4237
4238
4239
4240
4241
4242
4243
4244
4245
4246
4247
4248
4249
4250
4251
4252
4253
4254
4255
4256
4257
4258
4259
4260
4261
4262
4263
4264
4265
4266
4267
4268
4269
4270
4271
4272
4273
4274
4275
4276
4277
4278
4279
4280
4281
4282
4283
4284
4285
4286
4287
4288
4289
4290
4291
4292
4293
4294
4295
4296
4297
4298
4299
4300
4301
4302
4303
4304
4305
4306
4307
4308
4309
4310
4311
4312
4313
4314
4315
4316
4317
4318
4319
4320
4321
4322
4323
4324
4325
4326
4327
4328
4329
4330
4331
4332
4333
4334
4335
4336
4337
4338
4339
4340
4341
4342
4343
4344
4345
4346
4347
4348
4349
4350
4351
4352
4353
4354
4355
4356
4357
4358
4359
4360
4361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyAr |||||
4362
4363
4364
4365
4366
4367
4368
4369 CACCATCAAGACTCAACTTCATCTCTACGCCGCAAGAAC |||||
4370
4371
4372
4373
4374
4375
4376
4377 gProSerLysThrProSerProSerGlnProLysLysAsn |||||
4378
4379
4380
4381
4382
4383
4384
4385
4386
4387
4388
4389
4390
4391
4392
4393
4394
4395
4396
4397
4398
4399
4400
4401
4402
4403
4404
4405
4406
4407
4408
4409
4410
4411
4412
4413
4414
4415
4416
4417
4418
4419
4420
4421
4422
4423
4424
4425
4426
4427
4428
4429
4430
4431
4432
4433
4434
4435
4436
4437
4438
4439
4440
4441
4442
4443
4444
4445
4446
4447
4448
4449
4450
4451
4452
4453
4454
4455
4456
4457
4458
4459
4460
4461
4462
4463
4464
4465
4466
4467
4468
4469
4470
4471
4472
4473
4474
4475
4476
4477
4478
4479
4480
4481
4482
4483
4484
4485
4486
4487
4488
4489
4490
4491
4492
4493
4494
4495
4496
4497
4498
4499
4500

seq_name: /SIDS1/gcgcdata/geneseq/geneseq/NA1998.DAT:AAV18885

```
seq_documentation_block:
ID   AAV18885 standard; cDNA; 13815 BP.
```

AAV18885:

DT 17-AUG-1998 (first entry)

DE Mus musculus dystrophin cDNA.

PCR primer; phage vector; bacteriophage lambda; gene therapy;
antisense RNA; ribozyme; delivery; vaccine; antigen; treatment;
prevention; in vivo; in vitro; AIDS; cancer; cystic fibrosis;
acquired immune deficiency syndrome; Gaucher's disease; haemophilia;
post-translational modification; diagnostic reagents; production;
dystrophin; ss.

OS Mus musculus.

XX PN WO9805344-A1.

XX
PD
12-FEB-1998XX 03-III.-1997: 97WQ-11512928
PFXX
PR
11-MAR-1997. 97HC-0814859

PR 05-AUG-1996; 96US-0693865.
yy

PA (BGHM) BRIGHAM & WOMENS HOSPITAL.

PI Dubin DB, Kupper TS, Sarkar SN;

DR WPI; 1998-145345/13.

Improved gene therapy with phage vectors unable to release genetic material into cells - comprises surface ligand binding to cell-surface receptor, and transcribing nucleic acid in cell, useful for, e.g. treating AIDS

Example 9: pages 80-85: 110pp: English.

XX The sequence is that of murine full length dystrophin cDNA which
CC was used in construction of a chimeric bacteriophage lambda vector.
CC

Sequence 13815 BP: 4582 A: 2731 C: 3012 G: 3490 T: 0 other:

```
alignment_scores:
  Quality: 185.50
  Ratio: 0.278
  Percent Similarity: 43.284
  Gaps: 76
  Percent Identity: 18.559
  Length: 1541
```

alignment block:

US-09-512-581-2 x AAV18885

Align seq 1/1 to: AAV18885 from: 1 to: 13815

34 ArgLeuLysMetValIValIysThrPheMetAspMetAspGlnAspSerGl 50
:::>::: |||:::|||||:::|| |:::||||:
575 CAGGTCAAAATGTGATGAAAACATCATGGCTGGATTGCAGCAACCAG 624


```

823 MetValLysIleGlnAlaIleLysMetMetValArg..... 834
:::|||||::: ||| |||
2725 CCAATCAAGATTCAAGCGCTGTTGGAATGAAGAGAAATTTGAAATCGT 2774

835 .Trp.LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850
::: |||:::|||||:::|||||::: |||
2775 CTACAAGGTGTCGTTCACAGAACACGAGAAATCGATCATCTCCCTCACCA 2824
||| |||:::|||||::: |||
851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluI 867
||| |||:::|||||::: |||
2825 CTCGCCAGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 2857
||| |||:::|||||::: |||
867 nGlyIleSerLysProAspMetSerArgLeuArgLeuAlaGlyS 884
::: |||:::|||||::: |||
2858 .....TCTCAAGTT 2867

884 exAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
|| |||:::|||||::: |||
2868 CGAATACATTGAGGCTGAAGAGAACCG..... 2896

901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG 917
2896 ..... 2896

917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
::: |||:::|||||::: |||
2897 .....TCTGTTGTTCCACAATGCCAACAAGACTCA 2927

934 rGleuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysaspPro 950
2928 GG..... 2929

951 ValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
||| |||:::|||||::: |||
2930 ....AAATTCAGAGGTCTCACTTCCTCGGACTTCAA.....GTCTC 2969

967 nValArgArgGluTyrLeuLysGlnHisAla.....A 978
::: |||::: |||::: |||::: |||
2970 CACAGCGGAAACATCTCAACCAACACACTCTTGAAAACCTGGCTCTTG 3019

978 IValSerGluLysLeuLeuSerLeuLeuProGluTyr...ValValPro 993
|| |||:::|||||::: |||
3020 CCGAGCAAGACTTCGAGGTTCCTGTTGAGAACAGTACCTCCAGCCGAA 3069

994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
::: |||::: |||::: |||::: |||
3070 TTCACGCGCGCTCACCGTTGGACAACCTCGAGAGACCGAGCTCTCCCA 3119

1010 PileGluGlnLeuLysAspValLysGluCysLeuThrPheValLeuGluI 1027
::: |||::: |||::: |||
3120 AATCAAGTACAACAAGATCTTCGAAAG.....GAGTTCGAGCTCGAA. 3162

1027 leLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMet 1043
||| |||::: |||::: |||
3163 .....CAAGARACCCGAGAGCCGCGGTGAGTACTTCACCAAGATG 3204

1044 ValGluAsnIleLysGlnThrLysAsp..... 1052
|||::: |||::: |||
3205 GTCAAGAGCATCCAGAACAGCAAGGATACAAATCTGCTCTCCTTGAG 3254

1053 .AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValC 1069
::: |||::: |||::: |||
3255 ACTTGAGGCCCCACGTGACTACATCATGATGAACACCGAGGTACACCGTCT 3304

1069 ysasp.....ValalaMetAsn 1074
||| |||
3305 GTGACGAAGCAATCCGCATGTGCCAATGGAAAGTTGAAATTCGCGCTCC 3354

1075 IleIleMetSerLysSerThrThrTyrSerLeuGluSer..... 1087
|||::: |||::: |||
3355 CCNATCTCGAGGAACCAAGGAATGAGTCTCCGCTCCCACTTCTTGT 3404

```

```

1088 .....ProLysAspProValLeuProAlaArgPhePheThrGlnProA 1102
|||::: |||
3405 TGTCCGCCAGAGATCCCATCATCTCCGTCATCACTCCACGATCAACCAC 3454

1102 sLys. AsnPheSerAsnThrLysAsnTyrLeuPro..... 1113
::: |||::: |||::: |||
3455 ACCGTGAAGTTTC...ACTCTCTCTTACTCTACCTGGGGATCCGAGAG 3501

1114 .....ProGluMetLysSerPhePheThrProGlyLysProL 1126
|||::: |||::: |||
3502 AAGAGCGAAGTCAACGCTCAACCTTCAACAATCAACGAAGAACAGAA 3551

1126 ysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGly 1142
::: |||::: |||::: |||
3552 GAAATACGAGCGTAACTGACCGTCTC.....ACTTCAATGGAA 3589

1143 LysGlnSerGlnThrLysSerArgMetGluThrValSerAsnAlaSe 1159
||| |||::: |||
3590 TGCCAGAGTACGAGCTTTTGATCAAGGCCCGCTCTTAACGAGATCAAC 3639

1159 r.....SerSerSerAsnProSerSerProGlyArgIleLysGlyA 1173
|||::: |||::: |||
3640 GCTGTTGCTGAGTACAAGCTACCCGCTGAACCGAACAAGTTCTT..... 3684

1173 rGleuAspSerSerGluMetAspHisSerGluAsnGluAspTyr...Thr 1188
::: |||::: |||
3685 .....GCCGTTATTGTTGATCTGTTAAGCGCTTACAACTACTGGACT 3726

1189 MetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAs 1205
||| |||
3727 GTCTCGAGC.....CGCCAGAGAACACGAGAATGATCCGCT 3764

1205 pLeuValArgSerGluLeuGluLysProArgLysLysLysThrProV 1222
::: |||::: |||
3765 CGTCTGTTCAACTCACAGTTGAG...CCAATGTCGACAAATACGTCAACA 3811

1222 alThrGluGlnGlu.....GluLysLeuGlyMetAspLeuThrLys 1236
||| |||
3812 TCACCATCAATCCCCAATGAGCGGTTGAGCTCAAGAACGTT...CAA 3858

1237 LeuValGlnGluLysProLysGlySerGlnArgSerArgLysArgG 1253
|||::: |||::: |||
3859 GTGCCACGTGTGTACTTCCATCCATGCTCAACGTTCCGCTCAAG..... 3903

1253 yHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLys... 1268
|||::: |||::: |||
3904 .CACCTACTTAATGAGGTTCCGGATCTGTTTCAAGGTGCAAAAGAAC 3952

1269 .....ArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSer 1283
||| |||
3953 AATCCGTACTTTTCGATGATGCTCTCAAC..... 3984

1284 ProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGly 1300
3984 ..... 3984

1300 yThrPro.....LysGluG 1305
||| |||
3985 .ACTCCACTCACCACTGTTACTCTGATCGCAAGATTGCTCCGAGG 4033

1305 LuProThrMetLysThrSerLysLysGlySerLysLysSerGlyPro 1321
|||||::: |||
4034 AGCCAACTTTCGCTGTTCTCTCAAGAGACTGAAAGAAATTC..... 4077

1322 ProAlaProGluGlu.....GluGluGluG 1330
||| |||
4078 .....GAAGAGATGATCATCAAGGTTATCCGTGGAGAACAGAAAT 4118

1330 uGluArgGlnSerGlyAsnThrGluGlnLysSerLysGlnHisA 1347
||| |||
4119 TGTGCTCAACTCCAGAACGAGGAGATTCGCGTTAAGGTTACGGTAAGA 4168

1347 rgVal.....SerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360

```


1259 AGCGCGTCAACTTCTTAAGTCCATC.....CAAGAGACTCCATTC 1299
308 AlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAs 324
1300 CCATCTCAACCC..... 1311
324 pIleHisValProIleArgLeuGluCysValLysPheAla...SerHisC 340
1312ATTGCTGAAGCACTCATCAAGTTCCGCGAGTCTCGCG 1348
340 ysLeuMetAsnHisPro..... 345
1349 TTTCGAAGAACACAGGTTGTCCGACAATCTGCTTGGCTCGCGCGGA 1398
346 AspLeuAlaLysAspLeuThrGluTyrLeuLysValArg.....Se 359
1399 TCAGTTGTTCGCGGAATTTGTGACTACAAGATATCCGTCACCTCGTCG 1448
359 rHisAspProGluAlaIleArgHisAspValIleValSerIleValT 376
1449 CGAGGAACAACGCTGAGTCT.....AAGGAGAAGTTCTCCGCGTATTCA 1492
376 hrAlaAlaLysLysAsp.....IleLeuLeuVal 385
1493 TGCAACAATACAAAGGATGCTGAGACCACCTACGAGAAGATTTTGGCAGTC 1542
386 Asn.....AspHisLeuLeuAsnPheValArgG1 395
1543 AAGAGCATCGGAACGCTGGACTCGACATCTCCGTCAACAGCTCAACGA 1592
395 uArgThrLeuAspLysArg.....TrpArgValArgLysGluAlaMetM 410
1593 GATCATTGTTGACAAGCGCAATTAATCTTCAGTCCGCAAGGAAGCCATTG 1642
410 etGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGly 426
1643 ATGCCTCAGACTCTT..... 1659
427 LysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuHisI1 443
1660 AAGGACACCATGCCAGCG.....AAGATCCAAAAGGTACTCTCCCAAT 1703
443 eTyr.....TyrGlnAsnSerIleAspAspArgLeuLeuVal. 455
1704 CTACAGACAGACAATACGAGCGCAGAGATC.....CGCATGCTTGCTT 1747
456GluArgIlePheAlaGlnTyr 462
1748 TGTGGAGAATGATGCACACCCGCCAGAGAGTCTCTCTCGTCCCAAGTT 1797
463 MetValProHisAsnLeuGluThrGluArgMetLysCysLeuTyrTy 479
1798 GTTTCCTCAATGGAGAAAGAGACCAACCAAGTTGCTGCTCTTACCCA 1847
479 rLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMet 496
1848 C..... 1848
496 rLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeu 512
1849CAGATGATCCGTCACCTTGTCCCAAGTCCACCAACCCCTGT 1887
513 IleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVa 529
1888 TACCAACGT.....GTTGCTATCGTTTGTCTCAAAGGT 1919
529 lMetValIleThrArgAsnLeuProAspPro..... 539
1920 TCATTCTTCATCTGCTTATCAACCAAGAACAAATGATCGCTTCCTCCT 1969
540 ..GlyLysAlaGlnAspPheMetLysLys..... 548

1970 ACCTCAACTTCACACTCTCTTCAAAACTCTTCTCTGGAGCTCAATTT 2019
549 ...PheThrClnValLeuGluAspAspGluLysIleArgLysGlnLeuG1 564
2020 GACTTTGCCGCTATCTTCGAGAAGAACTCTTCTTCTAAAGGATCTTCA 2069
564 uValLeuValSerProThrCysSerLysGlnAlaGluGlyCysVala 581
2070 CGTCTCTCTTGATGCCGTCTTCGGA.....GGAACCTGGA 2104
581 rGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPhe 597
2105 ACAAGTACTTTGCCAAATCGGATTTCTCAGCAACACACATGGACAAATAC 2154
598 LeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAs 614
2155 GTTCAGATG.....GCTCTTGAAACTC..... 2178
614 pThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspG 631
2179GAAGCATT..... 2187
631 lyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArg 647
2188GAGAAGGAGTCCACAACCTGTTGTCCGCGGACGCCGTATCCAA 2229
648 AlaGlyLeuLeuLeuLysValLeuSerPhe.....Th 659
2230 ACCGGAATCAGCTTCTTAAGAACTTGCCCTGAAAATGAACATCCGTGC 2279
659 rHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaC 676
2280 TCGCCAGCAACTACAACGAGAAGGATGCTTCGCCATGGTCTACCTTC 2329
676 ysLeuLys...MetAsp.....AspGluLysValAla 685
2330 GTTACAAGGACATGGATTATGCCATCTCCCGTCGATACCCCACTCATT 2379
686 GluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAs 702
2380 GAGAAGTCAATTGAGAAGTACATCAGCAACGGA...AAAGTTGAG.... 2421
702 pPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerL 719
2422 .TTCAGGAAATCCGCGTCTCTT..... 2445
719 ysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAla 735
2446AACCAAGACACAGAGTTCGAGACTCAT.....CATGCC 2478
736 IlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuH1 752
2479GCTTACTTCTATGAGGTATCCG 2501
752 sLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuVal 769
2502 CAATTC.....CCAACCACTCTTGA.....CTTCCACTCATCG 2536
769 hrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrp 785
2537 TCTCTGGAAGATC.....CCAACGGTCTTTTCGGCTGAAGGA 2574
786 LysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspAr 802
2575 CAATTCTCTCTTGACTTGAGAAACCGAACTCGTTTGTACCGGTGAGGC 2624
802 gLeuProGlyLysLysThrThr.....LysLeuTrpValp 814
2625 TCGCCCATCGTTGCTGCCACCCACGCTTTACGAGATGAGAAATGTTACCC 2674
814 ro.....AspGluGluValSerProGluThr 822
2675 CACTCTTCGAGCAAGGAGTCAAGAGCGTTTCAATCCGTACCGCCTACACC 2724

[illegible]

[illegible]

826 eGlnAlaIleLysMetValArgTrpLeuLeuGlyMeTyLysAsnH 843
::: |||||:::
2859 TGAGCA.....GATATGAAGAATCAAG 2881

843 iSerLys.....SerGlyThrSerThrLeuArgLeuLeuThrIle 857
::::|||::: |||||:::
2882 AAGCTAAGTTACTAACGTCAATGACTTGGCTCGTCAAATTGCITTAATGTG 2931

858 LeuHis.....SerAspGlyAspLeuThrGluGlnGlyLysIleSer.. 871
:::||||::: |||||:::
2932 GAACATCCAAACTCCGAGCATCTTCCACGCTCAAACAAGCTCAAGCG 2981

872LysProAspMetSerArgLeuArgLeuA 881
|||||::: |||||:::
2982 TCGCTGGGCTCAACTTCGCCGACATGGTGCTCAATCAAGGAAGAACGACTCG 3031

881 laaLaGlySerAlaIleValLysLeuAlaGlnGluProCysTyHisGlu 897
::: |||||::: |||||:::
3032 AGAGAGCTCACCGGCTC...GAACATTCCGCATCGATTCCCAAGACA 3078

898 IleIleThrLeuGluGlnTyrlGlnLeuCys.....AlaLeuAlaIl 911
::: |||||::: ::|||:::
3079 GTTACTTGGATCGAGGACAAGACTCGTGACTCGAGATTTCGACGCGGT 3128

911 eaSnAspGluCysTyrlGlnValArgGlnValPheAlaGlnLysLeuHisL 928
::: |||||::: |||||:::
3129 GACCACCATCTTCCGGTGT....ATGAACCTTCAAA 3163

928 ysGlyLeuSerArgLeuArgLeuProLeuGluTyrlMetAlaIleCysAla 944
::: |||||::: |||||:::
3164 GAAGATTTCATGATGAACCGGATTTGGGA.....GCG 3198

945 LeuCysAlaLys.....AspProValLysGI 953
::: |||||::: |||||:::
3199 ATTCAAGCCAAGCTCGACTCGCTTCACAAAAGCCGATGACATCCGAAG 3248

953 uArgArgAlaHisAlaArGrGlnCysLeu.ValLysAsnIleAsnValArg 969
::: |||||::: |||||:::
3249 AGAAGCACCACAAGAGCTCAAGCCATTCTGTGAAG.....ATATCA 3289

970 ArgGluTyrlLeuLysGlnHisAlaAlaValSerGluLysLeuSer.. 985
|||||::: |||||::: |||||:::
3290 AGAGAATCCACCAAGCTCTGGGATATCTTTACAAAAGTTCGTGAAGCAC 3339

985 985

3340 GAAGCTAACTCGATGAAGCTGGAGATCTTCAACGCTTCCTCGTGATTT 3389

986LeuLeuProGluTyrlValValProTyrl 994
||| |||||::: |||||:::
3390 GGATCACTTCCAAGCTGGCTTACTGCCACCCAGACAGAGTTGCGCT... 3436

995 ThrIleHisLeuLeuAlaHisAspProAspTyrlValLysValGlnAspII 1011
::: |||||::: |||||::: |||||:::
3437CAGAGAAGAGCCCAATCTCT...TGGCTGAAGCTGACCAACTT 3477

1011 e.....GluGlnLeuLysA 1016
::: |||||::: |||||:::
3478 CTCATCAACACGCTGCCATCCGTCAAGAGATTGACCGATACGCTGAGGA 3527

1016 spValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
::: |||||::: |||||::: |||||:::
3528 TTACAAGAAGATCGCTGCAATGGGAGATCGGTGCTCACTCAAGATCAACACCG 3577

1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGI 1049
::: |||||::: |||||::: |||||:::
3578 ATCCACATACATGTTCTCTCGTCAACGCTTGTCTG..... 3613

1049 nThrLysAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeut 1066

3614GACTTCACAAGAGTGTGGGAAGAGTT 3638


```

103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAs 119
487 ATGCGT.....ATCCACTGCTTGGAAACGTTGAG... 516
119 pThrLysSerProGlnPheAsnArgTyrPheTyrLeu...LeuGluAsnI 135
517 ....AAAGTTTGCAATCTTCGCGCAATCAACACGTCATCTCGAAAACT 562
135 leala..... 136
563 TGGGTTACACCATATCGTTGACGGAAATTCACGCCCTCACGTTGGGTCTC 612
137 ...TrpVal.....LysSerTyrAsnIleCysPheGluLe 147
613 ATCTGGACCATCATCTTCGATTCAGATTCAGATCCAAAGATATACATTTGAAGA 662
147 uGluAspSerAsnGlu.....IlePheT 155
663 CGCTGACAAATCACGACACGACGTTCCAGCGAAGGAGGCCCTTGTACTGTGGT 712
155 hrGlnLeuTyrArgThrLeuPheSerValIleAsn..... 166
713 GCCAGATGAAACTGCTGGATATCCAAATGTCAACGTCGAAGAACTTCATCA 762
167 .....AsnGlyHisAsnGlnLysValHisMethIse 177
763 ACCAGTTGGAGAGATGGTCTTGCATTCATGCACTTATTCATAAAACATAG 812
177 tValAspLeuMetSerSerIleCysGluGlyAspThrValSerGlnG 194
813 ACCTGATCTCGTA.....GACTACGATAATTTACAAAAA. 846
194 luLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeu 210
847 .....TCGAATGCTCTTACAAATTTGCAATCTGCTTTTCATACCGCA 888
211 AsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGl 227
889 GAAATATCA...TTGGGATTTGGCCAAAGTTCCCTCGATGCCGGAAGATGTGAA 935
227 nAlaIleGluPro.....TyrIleThrThrPheP 237
936 CGTTGACCACCTGATGAGAAATCCATATTCATCTTATGTAGTTACATATT 985
237 heAsn.....GlnValLeuMetLeu 243
986 ATCATTACTTCAACAACTGAAACAGATAACATCCAAGGAAACGATATC 1035
244 GlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLe 260
1036 GGAAAGTG...ATTAAATGAGTTGATGGAGAAC.....GACAAATGAT 1076
260 uGluLeuTyr...AsnIleAspSerHisLeuLeu..... 270
1077 CAATCGATACGAAACTCTTTCATCTGATTTGCTGCAATGGATCAATGCCA 1126
271 ...LeuSerValLeuProGlnLeuGluPhe..... 279
1127 AGATCCAACCTCTCAACGAAACGTCATTTTGAGAATAATTTGGAAGGTGTT 1176
280 .....LysLeuTy 282
1177 CAAGACAGACTCACCAGGTTCACGACTACCGTACTCAGGAGAAACCCACC 1226
282 sSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysM 299
1227 AAGTTCGATGAAAGGGAGAGCTCGAAGTACTT.....C 1261
299 etPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeu... 314
1262 TGTTCACCTTTCATCAGTATGTCGTCGCCAACATCAGAGACCTTTCGPG 1311
315 .....TrpGlnCysTy 318

```

```

1312 CCGAGAGAAGGAAAACTGATGCTGATATCAACCGAGCATGGCAATCCCT 1361
318 rLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValL 335
1362 TCAGAAGGCAGAACACGACGAGAACTGCTCTGAAAGAGGAACTCATTC 1411
335 ys..... 335
1412 GTCAAGAGAAAACTCGAACAACTCGCAGCCCGGATTCAACAGAAAGCCGAG 1461
336 .....PheAlaSerHisCysLeuMetAsnHisProAs 346
1462 ATGCGAGAGACTTGGCTCACAGAAAACCAACGCGTGTGTCAGTCAGGACAA 1511
346 pLeuAlaLysAspLeuThrGluTyr...LeuLysValArgSerHisAspp 362
1512 TTTTCGAAACGACTTGTCTCGTGGAGAGCTGCCACCAAAAAAGCAC... 1557
362 roGluGluAlaIleArgHisAspValIle.....ValSer 373
1558 .....GAAGCCATCGAGACGGACATCTTGTCTAGCAGGAGCGTGTTCAA 1602
374 lleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHis..... 388
1603 GCCGTCGCTGCTGTGCTGGAGAGCTTGAAGCCGAGAACTACCATGATCA 1652
389 .....LeuLeuAsnPheV 393
1653 ACCCAAAATCAATGAGAGAAAGAGAAATGCTTCAACTCTGGAATATC 1702
393 aArgGluArgThrLeuAspLysArgTyrArgValArgLysGluAlaMet 409
1703 TTTTCCAACTTCTTTTGGCTCGC.....AGATTGCGATTGGAGTTGTCA 1746
410 MetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGl 426
1747 ATGGCCATTCAAAAGATCTC..... 1767
426 yLysAspAlaAlaLysGlnIleAlaTyrIleLysAspLysLeuLeuHisI 443
1768 .....CAGCATATGCTTCTTACTC 1786
443 leTyrTyrGlnAsnSerIleAspArgLeuLeuValGluArgIlePhe 459
1787 TCGATCTCATGGATGATATTAAGAGCAGACTTTTAAGCCGAGGATCTTGA 1836
460 AlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCy 476
1837 GCC.....CATCTTATGGAGCTAGAGATTTCCTTCAAAA 1871
476 sLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuA 493
1872 GCACCTCTTCTTGAGCTGATATTAATATTATTCGAGAACGAGTTAAAC 1921
493 snGluMetTrpLysCysGln..... 499
1922 ATTCTATTGCTCAAGCTCAACGCTCCGAAATCCAGATGGACCAGATGGA 1971
500 .....AsnLe 501
1972 AGTGATACAAGCCAGTTGAACCAAGAACCATTTGATGAGAGAGTGTATGT 2021
501 uLeuArgHisGlnValLysAspLeuLeuIleLysGlnProLysT 518
2022 TCTACAAAAACGTTACAGGAGCTTCTGATTGCTGCTGAACGTAAG. 2070
518 hrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg 534
2071 .....Cgt 2073
535 AsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGl 551

```



```

1185 GluAspTyrThrMetSerProLeuProGly.....LysLys 1197
      ||| : : : : : : : : : : : : : : : : : :
3697 GGACGACCACTAACCAATCTTTGGGTATCTCTATCAGGAAC 3746
      ||| : : : : : : : : : : : : : : : : : :
1197 sSerAspLysArgAspSerAspLeuValArgSerGluLeuLysP 1214
      : : : : : : : : : : : : : : : : : : : :
3747 TGCTTCATTGGTGAATAACAGCGCTTC...TTAGAGATACAAGAT 3793
      ||| : : : : : : : : : : : : : : : : : :
1214 roArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGly 1230
      : : : ||| : : : : : : : : : : : : : : :
3794 TGCAATTGGAATTAAGCTATTGAGGAGCAATAACCAAAAGCTTGCA 3843
      ||| : : : : : : : : : : : : : : : : : :
1231 MetAspAspLeuThrLysLeuValGlnGluLysProLysGlySerG 1247
      ||| : : : : : : : : : : : : : : : : : :
3844 ATGGAGAATA.....AAACTTCCTCTGCCCG 3872
      ||| : : : : : : : : : : : : : : : : : :
1247 nArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnP 1264
      : ||| : : : : : : : : : : : : : : : : :
3873 AAGAAGCAGAAGAGAAATGGAAGCCTCCGGAACAGAACATCAATCT 3922
      ||| : : : : : : : : : : : : : : : : : :
1264 rProGluGluLysArgLeu..... 1270
      : : ||||| : : : : : : : : : : : : : :
3923 ACTCCGAGGAACCGAGATGATCACAAAAGATATCGTACTCGATCAGGT 3972
      ||| : : : : : : : : : : : : : : : : : :
1271 .....LysGluAspLeuGluAsnG 1278
      : : ||||| : : : : : : : : : : : : : :
3973 TCATGTCATCATATGGGATAAGTACGAGAGACATTTTGAAGATAGA 4022
      : : ||||| : : : : : : : : : : : : : :
1278 uAspGluGlnAsnSerProProLysLysGlyArgGlyArgProP 1295
      : ||| : : : : : : : : : : : : : : : : :
4023 AGATGATCACAGT.....TTAGAGGCAAGTCACAAAATCCTCCGA 4063
      : : : : : : : : : : : : : : : : : :
1295 ysProLeuGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
      ||| : : : : : : : : : : : : : : : : : :
4064 AA.....GATAAATCCTTGTCTGAAGATCATTGTTGTGGAC 4101
      : : : : : : : : : : : : : : : : : :
1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluG 1328
      ||| : : : : : : : : : : : : : : : : : :
4102 AAATGAGATCTCCGATGATTCAGGATCCAAACAGATCAACAA 4151
      : : : : : : : : : : : : : : : : : :
1328 uGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSer 1345
      : : : : : : : : : : : : : : : : : :
4152 GAGAAAGTTCTCGAGAGACTTAATTCAGATTACAGAAAGCTCTCAAATC 4201
      : : : : : : : : : : : : : : : : : :
1345 lnHisArgValSerArgAlaGlnGlnArgAlaGluSerProGlu.Se 1361
      ||| : : : : : : : : : : : : : : : : : :
4202 TTCATGTGCGCGTTGAAGATTGAAAATCAAGTGGAGACAGAGAA 4251
      : : : : : : : : : : : : : : : : : :
1361 rSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgP 1378
      ||| : : : : : : : : : : : : : : : : : :
4252 GACGAGAAAGAAAGAAACGAGTACG...AGACGATCAAGGACAGAT 4298
      : : : : : : : : : : : : : : : : : :
1378 roSerLys 1380
      : : : : : : : : : : : : : : : : : :
4299 CAACGAAG 4306
      : : : : : : : : : : : : : : : : : :

```

seq_name: /SDS1/gcgdata/geneseq/geneseq/NA2001.DAT: AAC90918

seq_documentation_block:

ID AAC90918 standard; cDNA; 6981 BP.

AC AAC90918;

XX AAC90918;

DT 19-MAR-2001. (first entry)

XX C. elegans spectrin beta-chain/fodrin cDNA sequence SEQ ID NO:11.

DE Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
KW protein-protein interaction; identification; ss.

XX Caenorhabditis elegans.

OS

```

PN WO200073328-A2.
XX
PD 07-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-EP05108.
XX
PR 01-JUN-1999; 99GB-0012755.
XX
PA (DEVG-) DEVGEN NV.
XX
PI Van Crieckinge W, Roelens I, Bogaert T, Verwaerde P;
XX WPI; 2001-016508/02.
XX P-PSDB; AAB50652.
XX
PT Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
PT a human unc-5Hs1 cDNA, useful in yeast two hybrid experiments for
PT identifying unknown human cDNAs which encode proteins that interact
PT with the human unc-5C protein -
XX
PS Example 4; Page 108-110; 246pp; English.
XX
CC The present invention describes 3 variants of human unc-5C cDNAs
CC (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
CC unc-5C transcripts, and a human unc-5Hs1 cDNA which shares homology with
CC the Rattus norvegicus unc-5Hs1 cDNA. Also described are assays based on
CC protein-protein-interactions between the unc-5 protein and a variety of
CC different interacting proteins. The unc-5C variant cDNAs and unc-5Hs1
CC cDNA are useful in methods for identifying compounds which reduce or
CC inhibit the lethal phenotype associated with the expression of the
CC unc-5 death domain in yeast. They are also useful in yeast two hybrid
CC experiments for identifying unknown human cDNAs which encode proteins
CC that interact with the human unc-5C protein. AAC90914 to AAC90971 and
CC AAB50646 to AAB50693 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 6981 BP; 2165 A; 1605 C; 1641 G; 1570 T; 0 other;

```

alignment_scores:

Quality: 186.50 Length: 1680
Ratio: 0.256 Gaps: 78
Percent Similarity: 43.333 Percent Identity: 17.917

alignment_block:

US-09-512-581-2 x AAC90918 ..

Align seg 1/1 to: AAC90918 from: 1 to: 6981

```

17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValAr 33
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 CCAACTAGTCTGAGTTCTGATTCTAGTATGCGACAAAATCTATAAA 281
      : : : : : : : : : : : : : : : : : :
33 gArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerG 50
      : : : : : : : : : : : : : : : : : :
282 TTCAGTTGACTTGAGGTTTCAGAGGATTGAGAGAT..... 315
      : : : : : : : : : : : : : : : : : :
50 luGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
      : : : : : : : : : : : : : : : : : :
316 .....GAACGTGAACCTC.....GTGCAAAAGAAA 339
      : : : : : : : : : : : : : : : : : :
67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuAlaCysCy 83
      : : : : : : : : : : : : : : : : : :
340 ACATTTACAAAATGGGTCAACTCACACTTG...GTTTCAGTCAGCTGCAA 386
      : : : : : : : : : : : : : : : : : :
83 sLeuAlaAspIlePhe.....ArgI 90
      : : : : : : : : : : : : : : : : : :
387 AGTACAAGACCTCTACATGGATATGCGAGACGGAATAATGCTCTCCGAC 436
      : : : : : : : : : : : : : : : : : :
90 letYrAla.....ProGluAlaProTyrThrSerProAspLys 102
      : : : : : : : : : : : : : : : : : :
437 TCCTCGCAGTTCTTTCGCGTGAACGTCTGCCGAAACGACCGCGGAAAA 486
      : : : : : : : : : : : : : : : : : :

```


[illegible]

420 laLeuGlnSerAlaAalagLylyAspaAlaIalysGlnlleAlatrpille 436
:::|||||:::
1328 GTTTCACATCCAGAAAGCTTTGGGTCTCTCTCAAGGAGTTACAGGAGGAG 1377
:::|||||:::
437 LysAspLysLeuHisiIleTyTYrGlnAsnSeriLeaspaArgLe 453
:::|||||:::
1378 AACCTCAAAGCTTAAGGAATATCCGGGAAGAAGAAGATTGTAAAAGACGGC 1427
:::|||||:::
453 uLeuValGluArgIlePheAlaGlnTyMetValProHisAsnLeuGluTy 470
|||:::|||||:::
1428 CCTGATTGAGAAG.....TTAGAAA 1447
:::|||||:::
470 hrThrGluArgMet.....LysCysLeuTyTYrLeuTyAlaThrLeu 484
|||:::|||||:::
1448 TGATGGAAAAAGCTTGTTTCAGAAAAAACCTTCTCTGGAGAACTCCATCATCA 1497
:::|||||:::
485 AspLeuAsnAla.....VallysAlaLeuAs 493
|||:::|||||:::
1498 GATTTCGAATGCTGAGCTGGAAACAATCAGAGGAAGCTGAAACATTTGGA 1547
:::|||||:::
493 nGluMetTrpLysCysGlnAsnLeuLeuArg.....HisG 505
|||:::|||||:::
1548 GGAA.....GCTTCTATGTCACCTGCTGAAGAGAAATCAGGCCTTCATT 1591
:::|||||:::
505 lnValLysAspLeuLeuAspLeuLysGlnProLysThrAspAlaSer 521
:::|||||:::
1592 CTGAGAAGGACATCTTGATTTCGCCGTGCAGAGTGTCTACAGAAAAACAGC 1641
:::|||||:::
522 ValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAs 538
|||:::|||||:::
1642 ...AAGAAACTCTCTGAGGAAAAACATGGTACTCGAGAATTTCTCTTTTAA 1688
:::|||||:::
538 pProGLyLSAlaGlnAspPheMetLysPheThrGlnValIleGluLeu 555
:::|||||:::
1689 TGCCAATGTAGAAGCTTGAAGAACTCAAGTGCAAATTGAAAGCCTAGAAG 1738
:::|||||:::
555 sp.....AspGluLys..... 558
:::|||||:::
1739 AATCATGCCACCTGCTCAATGATGATGACAAACCACCTTTAACTAGTGAAGA 1788
:::|||||:::
559IleArgLysGlnLeuGluValle 566
:::|||||:::
1789 GAAAGCCTACTCTCATATTGATACAAATGAGAAAAAGAAATTGGAAGTCT 1838
:::|||||:::
566 uValserProthrCysSerCysLys.....G 575
|||:::|||||:::
1839 GCAGAAAGACATCGCAAAATAAAGTGAAGTTTGGAAATTAGCCACTG 1888
:::|||||:::
575 lnAlaGluGlyCytsValArgIleThrLysLysLeuGly..... 588
:::|||||:::
1889 AGAGGGAGTCTTCTTTCAGAAAAATT...GAAGAGTTGGGAGTGTCTTTA 1935
:::|||||:::
589 AsnProLysGlnProThrAsnPropheLeuGluMetIleLysPheLeu 605
|||:::|||||:::
1936 AATGCCAAGGAC.....TGTGAGTATGCTAGTGTGTCGAATTTTCAGA 1979
:::|||||:::
605 uGluArgIle.....AlaProValHisIleAspThrGluSerI 618
|||:::|||||:::
1980 GAGTCGGATGAATGGCATGGAAATCAAGCATCCATCACCTTCAGATGAAGA 2029
:::|||||:::
618 leSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAsp 634
|||:::|||||:::
2030 ATCAATCCGGGTGAGAGAATATCAAGTGGAACTGGATCGAGCTCATGAT 2079
:::|||||:::
635 AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGI 651
|||:::|||||:::
2080 GCTCACATTGAGATTATTGTTCTCGAAAAATGTTTGAGGACTGGCTT.. 2127
:::|||||:::
651 uLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaG 668
|||:::|||||:::
2128G 2128
|||:::|||||:::
668 luThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysVal 684


```
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 188.50      Length: 1620
    Ratio: 0.252        Gaps: 78
Percent Similarity: 46.111 Percent Identity: 18.951

alignment_block:
US-09-512-581-2 x AAC45991 ..

Align seg 1/1 to: AAC45991 from: 1 to: 4689

22 IleSerAspLysIleSerLys.....GluGluMetVa 32
||||| ||| |||||
58 ATAAGCCCCAAGAACTCGAAATGGCTTCAGGAGAACTTACAGATATGGA 107
||||| ||| |||||
32 lArgArgLeuLysMetValValLys.....ThrPheMeta 44
||||| ||| |||||
108 TAGTAAGTCAAGCAAAATGATCAAGCTGAAAGGAAAGTAGCTTAGCTC 157
||||| ||| |||||
44 sPMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
||||| ||| |||||
158 AGTTTCATCAGAAATTTGGAGAACTGCTTAATTG..... 192
||||| ||| |||||
61 LeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspValar 77
||||| ||| |||||
193 .....GAGTCTGAGGTATCTCGTGCACAAAGAGGACTCTAG 227
||||| ||| |||||
77 gLeuLeuValAlaCysCysLeuAlaAspilePheArgIleTyrAlaProG 94
||||| ||| |||||
228 GGTGCTTATTGAA.....CGAGCCACAAGAGCCG 256
||||| ||| |||||
94 luAlaProTyrThrSerProAspLysLeuLysAspilePheMetPheIle 110
||||| ||| |||||
257 AAGCTGAGGTGCAACG.....CTC 276
||||| ||| |||||
111 ThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnAr 127
||||| ||| |||||
277 AGAGAATCCCTCAGCAAAAGTAGAGTTGAAAGGAGTCCAGCCTTCTTCA 326
||||| ||| |||||
127 gTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleC 144
||||| ||| |||||
327 GTATCAGCAATGCTGCCAAACATAGCT..... 354
||||| ||| |||||
144 ysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
||||| ||| |||||
355 ....GATTTGGAGGTGCAATTTCTCTAGCCCAAGAGGAGCTGGAGAA 399
||||| ||| |||||
161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMe 177
||||| ||| |||||
400 GTAGATGAGCGAGCCCAACAGCCGAAAGCTGAAACTCTAGCACTGAAGCA 449
||||| ||| |||||
177 tValAspLeuMetSerSerIleIleCysGluGlyAspThrValSer...G 193
||||| ||| |||||
450 AAGCCTTGTAGTTCCGAAACTGACAAAGGAGCTGCTTGTTCATATC 499
```



```
3651 ..... |||||
1277 nLuAspGluGlnAsnSerProLysLysGlyLysArgGlyArgProp 1294
3668 TGAGTTTGGCCAGCAGGACTTCAGAAAGTTGTAAGAAAGGTTTGCTG 3717
1294 roLysProLeuGlyGlyThrPro.....LysGluGluProThrMet 1308
3718 ACATCCCGACGAGAAAGACTAGCCCATATATCTCGCGAAGAACCAACCATG 3767
1309 LysThrSerLysLys.....GlySerLysLysLysSerGlyProPr 1322
3768 GCACTGGAGAGCGCCCGCTGGCTGCACACAGTTACCCCTATCCCC 3817
1322 oAlaProGluGluGluGluGluArgGlnSerGlyAsnThrGluG 1339
3818 ACTGACTGTCCCAACAAATCTTCGAGAGTCTCTCAAAACCAACAGCTG 3867
1339 LysSerLysSerLysGlnHisArgValSerArgArgAlaGlnArg 1355
3868 GTGGCAGCAGATCAAAAGGTGAAGTGTCTCAGCGG..... 3905
1356 AlaGluSerProGluSerSerAlaIle.....GluSerThrG 1368
3906 .....AGCCAGTAGATTACAGGCACCATCTCGAGAACCCACACGAA 3949
1368 nSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerPro 1385
3950 ATCCGTCCCACTCAATATCTTCGAGAGAGTCCGACTGACAGGCCCA 3999
1385 erGln 1386
4000 GAGAG 4004

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC45991
seq_documentation_block:
ID AAC45991 standard; DNA; 4689 BP.
XX AAC45991;
XX
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 48513.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
```



```

826 CGCAGCCATAGCCTGGCAAGAGCCGACGCAAGCAAGCAAGCAAGCTGA 875
1235 LysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLysA 1252
876 AGAGAAGATCCAAACAAATGACAAATCGGGAACCAAGAGCCGGAGTC 925
1252 rGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLys 1268
926 CTAGCAGGCATAAAGTAAGAGCAAAAGTCGGAGCAGGAGTCCAGGAGG 975
1269 ArgLeuLysGluAspLeuGluAsnGluAspGluGlnAsnSerProPr 1285
976 AGAGTGGAGGAGG..... 990
1285 oLysLysGly.....LysArgGlyArgProProLysProLeuGlyGly 1300
991 ..AAGCAGGAGGTGTGAGCAGGGCAGG..... 1017
1300 lyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLys 1316
1018 .....AGCCAGGAGAGAGCCTCCGCCAGAGTCGGAGCGGAGCAGG 1059
1317 LysLysSerGlyProAlaProGluGlu..... 1327
1060 AGCAAGCGGGCAGCAGAGCGGAGCAGGAGGAGCGCCCAAGCAAGCAAGG 1109
1328 GluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSer 1343
1110 ACAAGAGGAAGAGCAGGAGAGAGCAGAGAGCAGGCGCAGTCGCAGT 1159
1344 LysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProG 1360
1160 CCGAGCCGAGCAAGAGTGTGAGAGGAGCAGAAAGCGGAGCAGCAAGCG 1209
1360 uSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGly 1377
1210 CAGCAAGCGRGAGCAGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1259
1377 rgProSerLysThrProSerProSerGlnProLysLys 1389
1260 CCGAGTCCAGATCTCCATCCCGTCGTCGTCAAGGAG 1297

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ67602

seq_documentation_block:

ID AAQ67602 standard; cDNA; 4868 BP.

XX AC AAQ67602;

XX DT 20-NOV-1994 (first entry)

XX DE Retinoblastoma protein Ap10.

XX KW Retinoblastoma-associated polypeptide; RAP; Ap12; Ap10;

XX KW transcription factor; EF2; cell cycle; ss.

XX OS Homo sapiens.

XX PN W09412521-A.

XX PD 09-JUN-1994.

XX PF 19-NOV-1993; 93WO-US11310.

XX PR 20-NOV-1992; 92US-0979156.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Lee W, Shan B;

XX DR WPI; 1994-200185/24.

XX

PT Nucleic acid encoding retinoblastoma-associated polypeptide(s) -
PT used for producing prods. for use in studying cell cyclng and
PS dysregulated cell growth.
XX
XX PS Disclosure; Page 38-40; 77pp; English.
XX
XX CC cDNA encoding a retinoblastoma (RB) associated protein (RAP), Ap12,
CC was cloned by direct screening of cDNA expression libraries using
CC purified RB protein as probe. Ap12 showed transcription factor E2F
CC and RB binding activities. The 5' and 3' cDNA sequences were
CC determined for Ap2 (AAQ67395, AAQ67596), Ap8 (AAQ67597, AAQ67598) and
CC Ap15 (AAQ67599, AAQ67600), and full-length encoding sequences for Ap4
CC (AAQ67601) and Ap10 (AAQ67601), which also encoded RB-associated
XX proteins.

SQ Sequence 4868 BP; 1755 A; 837 C; 1165 G; 1111 T; 0 other;

alignment_scores:

Quality: 189.50 Length: 1535
Ratio: 0.275 Gaps: 64
Percent Similarity: 44.886 Percent Identity: 19.349

alignment_block:

US-09-512-581-2 x AAQ67602 ..

Align seg 1/1 to: AAQ67602 from: 1 to: 4868

```

19 ValLysGluIleSerAspLysIle.....Se 27
156 ATGAAGAAGATTAGACTCAAACTCCATTACAGGAGGTACAACCTAATGAC 205
27 rLysGluGluMetValArgLysMetValLysThrPheMetA 44
206 CAAATTTGAAGCATGCATAGAAATGGAAAAATAGTTGGGAACTTAAGA 255
44 sMetAspGlnAspSerGluGluLysGluLeuTyr..... 56
256 AAGAAGAACTCAGATTTAAGTGAATAATTTCTTGTGATCAC 305
57 .....LeuAsnLeuAlaLe 61
306 CAGGAGTTACTCCAGAGATGAACTCTTGAGGCTCAATTCGTGATTT 355
61 uHisLeuAlaSerAspPheLeuLysHisProGlyLysAspValArgL 78
356 AGAATATGCATGCAGAT.....AAATCATCAGCTGAAGATATTGGAG 396
78 euLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
397 ATAAATGTGGCC..... 407
95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIlePh 111
408 .....AAGTGAATGCAGCTGGAGAGAGATTTCTTGTGATGGGA 448
111 rArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgT 128
449 AAATCAGCTGAGTAGATCAGATCGGAGAAAGCTAGCATTGAGCATGAAG 498
128 yrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCys 144
499 CCTCTACTCTGGAGGCTGACTTAGAGGTAGTTCAACACAGAGAGATGT 548
145 PheGluLeuGluAspSerAsn.....GluIlePheThrGlnLeuTyrAr 159
549 TTAGAAAAGACAAATGAAATAGCAGAGGTTATTGTGCTCTGAAGA 598
159 gThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMeth 176
599 AGAACTC...TCAGTGGTCAAGTGAGAGAAACCAGCTT.....CGTG 639
176 isMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSer 192

```



```

1277 TCATGG.....ATCAATCTCAACAAGTATT.....GCTGA 7308
1279 pGluGlnAsnSerProProLysLys.GlyLysArgGlyArgProProLys 1295
7309 CACAGAGACGCTTGAGGCTGAGGCTGAGGAGCGGTGCAGGCTCTCGCA 7358
1296 ProLeuGlyGlyThrProLysGluGluProThrMetLysThrSerLy 1312
7359 GAGATC.....TGGAAACTTCCTG 7378
1312 sLysGlySerLysLysLys 1318
7379 AAGTGGATCAAGAGCAG 7397

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC98914

seq_documentation_block:
ID AAC98914 standard; cDNA; 2212 BP.

```

AC AAC98914;
DT 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:142.
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX Homo sapiens.
OS
XX
XX WO200055320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-579444/54.
XX
XX P-PSDB; AAB54149.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 1; Page 600-601; 1379pp; English.
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX proteins, called pancreatic cancer antigens, given in AAB54008 to
XX AAB54466. The human pancreatic cancer antigens have cytostatic,
XX neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
XX gynaecological, cardiant and antiinflammatory activities, and can be used
XX in gene therapy. The polynucleotide and proteins can be used for
XX preventing, treating, or ameliorating a medical condition or in assays
XX for diagnosing a pathological condition or a susceptibility to one in a
XX subject. Binding partners to the proteins and the activity of the
XX proteins can be identified. The pancreatic cancer antigens can be used to
XX detect, treat or prevent pancreatic disorders, especially cancer.
XX Agonists and antagonists to the antigens can be screened for. The
XX pancreatic cancer antigen polynucleotides can be used to design nucleic
XX acid hybridisation probes that can be used in chromosome mapping, linkage
XX analysis, tissue identification and/or typing and a variety of forensic
XX and diagnostic methods. The proteins can be used to generate antibodies
XX

```

CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 2212 BP; 648 A; 488 C; 619 G; 456 T; 1 other;

alignment_scores:

Quality: 190.50 Length: 398
Ratio: 0.907 Gaps: 18
Percent Similarity: 52.764 Percent Identity: 24.372

alignment_block:

US-09-512-581-2 x AAC98914 ..
Align seg 1/1 to: AAC98914 from: 1 to: 2212

```

1041 ArgLysMetValGluAsnLleLysGlnThrLysAspAlaGlnGlyPro... 1056
1042 :||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 CGAGTAATTGTTGAGCAT.....GCCCGCGGCCACG 276
1057 .....AspAspAlaLysMetAsnGluLysLeuTyrThr 1068
1058 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
277 GCGAGATGCCAGTTACGGTCTCGACGCATGATATGCTTATAGAGAAG 326
1068 alcysaspvalAlaMetAsnLleMetSerLysSerThr..ThrTyr.. 1083
1069 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 TGCGCGAGATAAATATGCGCCTCTACTCGCAGAGTACAGACTTATTG 376
1084 .....SerLeuGluSerPro.LysAspProVal 1092
1085 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
377 TGGAGAAATTTGTCAGTCGGTGCAGTGCAGACCTAAAGATTATATG 426
1093 LeuProAlaArgPhePheThrGlnProAsp.....LysAs 1104
1094 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
427 CGTCAGCAGGAGAGTACTTATGCGAGTCTCACAAGGAGCGCAAAA 476
1104 n.....PheSerAsnThrLysAsnTyrLeuP 1113
1105 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 TGAAGGGGTGATTGAATTTGTATCTTATCTGATATGAAAAGAGCTTGG 526
1113 roProGluMetLysSerPhePheThrProGlyLysProLysThrAsn 1129
1114 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
527 AA.....AAGTTGGATGGAACGTAA 546
1130 ValLeuGly.....AlaValAsnLysProLeuSerSerAl 1141
1131 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
547 GTCATATGGGAGAAAATCAGATTAGTTGAAGACAGCCA..... 585
1141 aGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsn 1158
1142 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
586 .GTTCCAGACGACGCGGCTCTACTCCAGAACGCGGAGTCATTCAAGGT 634
1158 laSerSerSerAsnProSerSerProGlyArgLleLysGlyArgLeu 1174
1159 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 CTCGCTCTCGAAGCAGACATTCCTCCGTAAGACAGAGA.....AGCCGAAGT 678
1175 AspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSer 1191
1176 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
679 GCGCAGCAGAAAAGCAGTCAATCTTAAGAGTAGATCTCGGTCCAGGTGG 728
1191 rProLeuProGlyLysSerAspLysArgAspSerAspLeuVala 1208
1192 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
729 CTCCCGCTCCCGAGCAAGACGCGGAGCGGAGCCAGAGT...CGGAGCC 775
1208 rGserGluLeuGluLysProArgGlyArgLysThrProValThrGlu 1224
1209 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
776 GGAGCAAGAAAGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 825
1225 GlnGluGluLysLeuGly.....MetAspLeuTh 1235

```


4147 ATCCACGAGAGATCTCAGCCCATGAGCTACCCCTAGAGGAGTTGAGAA 4196
 297 laLysMetPheGlyAlaLysAspSerGluLeuAlaSer.GlnAsnLysPr 313
 4197 GAAATATG.....CGTCTCAGCCCTGACCTCCCGAGAGAGTAGG 4237
 313 oLeu.....TrpGlnCysTyrLeuGlyArgPhe.Asn 323
 4238 ACTGCCAGAGGAGGAAGTCAGATGGATGTGCTACAGAGAAACTCCGAGA 4287
 324 ArgPheHis...ValPro.Ile..... 329
 4288 GTGTGCCCAAGATTCACAGCTTTTCCAGAACCCAGCTAACTTCGAGCAGC 4337
 330 ..ArgLeuGluCys.....ValLysPheAlaSerHis 339
 4338 GCATGCTGGACTGCAAGCGTGTGCTGGATGGCGTGAAGACGAACTTCAC 4387
 340 CysLeu.....MetAsnHisProAs 346
 4388 GTTCTGGATGTGAAGGACGTAGACCCCTGACGTCATACAGACGCACTTGA 4437
 346 p.....LeuAlaLysAspLeuThrGluTyrLeuLysValArgS 359
 4438 CAAGTGTATGAACCTGTATAAAGCTTTGAGTGA..GTCAAACCTTGAAG 4484
 359 erHisAspProGluAlaLysArgHisAspVal..... 370
 4485 TGGAACTGTGATTAAACAGGACGACATATGTCACGAACACAGCAACG 4534
 371IleValSerIle..... 374
 4535 GACACCCCAAGGATGGATGAGCAGCTGACTTCCCTGAAGGTTCTTTA 4584
 375ValThrAlaLysLysAspLeuLeuVal 385
 4585 CAATGACCTGGCGCACAGGTGACAGAGGAAACAGGATCTGGAAGAG 4634
 385 alAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArg 401
 4635 CATCACAGTTGCCCGGAAATGAAGAAGAGCGTCTCTCTCTGAA 4684
 402 TPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLy 418
 4685 TGGCTTTCTGCTACTGAA.....ACTGAATTGGTACAGAA 4719
 418 sTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaLysGlnIleAlat 435
 4720 GTCCACCTTCAGAAAGTCTGCTGGT...GACTTGGATACAGAAATTTCT 4766
 435 rpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAsp 451
 4767 GGGCTAAATAATCTTGAGGATCTGGAAGACAGAAAGCTGATTTAAAT 4816
 452 ArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLe 468
 4817 ACCATCACAGAGCTAGTGTCTGCCCTGCAAACTGTATGGGCGAGTGA 4866
 468 uGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuA 485
 4867 GCCTATTTTGAAGAGAGCTCTGGCTC..... 4894
 485 sPLeuAsnAla.....ValLysAlaLeuAsnGluMetTrpLys 497
 4895 ..CTTAACGCTGGGTGGAGCCGAGTTCTGCTACCTGGACTGAAGATTGG... 4939
 498 CysGlnAsnLeuLeuArgHis..... 504
 4940 TCCATACCTTGAATCAACCATCAGAACCAAGCTAGAAATATTTTATGGGAA 4989
 505GlnValLysAspLeuLeuAspL 512
 512 519


```

6920   ...  |||::|||  :::::  :::  |||::
      .AGCAGAACCAGAAATACAGGAGCTCATCGAAAAAATCTCCGCTGT  6968

922   heAlaGlnLysLeuHisLysGlyLeuSerArgLeuProLeuGlu  938
      :::::  |||  :::::  |||  |||::
      .ATTCAAGCAA...CACACTTGCTCAGCAGTCTCTCTAGTAGCTCAA  7015

939   TyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAr  955
      |||::|  :::::  :::::  :::::  :::::  :::::
      .....AAGGAACAATGA  7026

955   gAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgGluTr  972
      :||||  :::::  |||::  :::::  :::::  :::::  :::::
      .....TTAAATATAAGGAATCTC  7064

7027   AGCACAC...AAACATTGCATG.....  7164
      :::::  :::::  :::::  :::::  :::::  :::::  :::::
      .....GlnHisAla  977

972   yrLeuLys.....  7114
      ::  :::::  :::::  :::::  :::::  :::::  :::::
      TGGCATGCCAGCTCTCCAGATCCCTTGGCAGCTTGCAAACTGAGCATGTT  7114

978   AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTy  994
      :::::  :::::  :::::  :::::  :::::  :::::  :::::
      ..AGCTAAATACTCAACTCAGACCCCTCTGAACAAAATTTAAGGTGTGATA  7164

7115   AAGCTAAATACTCAACTCAGACCCCTCTGAACAAAATTTAAGGTGTGATA  7164

994   rThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspI  1011
      ::  :::::  :::::  :::::  :::::  :::::  :::::
      CCGAACTGCTGCAGTCAAGAAGATCATAGCTTCATCAAA....GATT  7208

7165   CCGAACTGCTGCAGTCAAGAAGATCATAGCTTCATCAAA....GATT  7208

1011   leGluGlnLysAspValLysGluCysLeuTyrPheValLeuGluIle  1027
      |||  |||::|  :::::  :::::  :::::  :::::  :::::
      .ATGAG.....AAGGACCTTCTGCT  7209

7209   ATGAG.....AAGGACCTTCTGCT  7228

1028   LeuMetAlaLysAsnGlnAsnSerHisAlaPheIleArgLysMetVa  1044
      |||::|  :::::  |||  :::::  :::::  :::::  :::::
      .....GAGCAAAAGAGGCGATGATGAGCTGGGACTCCAAC  7263

7229   .....GAGCAAAAGAGGCGATGATGAGCTGGGACTCCAAC  7263

1044   lGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys.  1060
      :::::  :::::  |||  :::::  :::::  :::::  :::::
      GCAGTGTGAGCAG..CAGCGCAGAAAATGTGCGATTCGCATCTG  7310

7264   GCAGTGTGAGCAG..CAGCGCAGAAAATGTGCGATTCGCATCTG  7310

1061   .....MetAsnGluLysLeuTyrThr  1067
      .....  |||::|  :::::  |||::|  :::::  :::::
      AGAACTCAAGTCTGCGAAATGAAATCTTGAATGAGTACTTTTTAAA  7360

7311   AGAACTCAAGTCTGCGAAATGAAATCTTGAATGAGTACTTTTTAAA  7360

1068   ValCysAspValAlaMetAsnIleMetSerLysSerThrTyrSe  1084
      |||::|  :::::  |||  :::::  :::::  :::::  :::::
      AAAGCAAAATATAATTACAGAGTGTCCAGGATGACTTTTC  7398

7361   .....AAAGCAAAATATAATTACAGAGTGTCCAGGATGACTTTTC  7398

1084   r.....  1085
      |  :::::  :::::  :::::  :::::  :::::  :::::
      AGAGGTGCAGGTATTCTTAATCAAGTAGGATCAACACTGCAAGAAGAGC  7448

7399   AGAGGTGCAGGTATTCTTAATCAAGTAGGATCAACACTGCAAGAAGAGC  7448

1085   euGluSerProLys.....  1089
      |||  |||  :::::  :::::  :::::  :::::  :::::
      TTGACCAACAAGAACGGCTTTATGCGAGTGGTGGAGGAATTTGGAGATCTG  7498

7449   TTGACCAACAAGAACGGCTTTATGCGAGTGGTGGAGGAATTTGGAGATCTG  7498

1090   ....AspProValLeuProAlaArgPheThrGlnProAspLysAs  1104
      |||  :::::  :::::  :::::  |||  :::::  :::::
      CAGCTCGATGCTAAGAACTCAGTCAAGGCATGCAACAGGAAATAGGCG  7548

7499   CAGCTCGATGCTAAGAACTCAGTCAAGGCATGCAACAGGAAATAGGCG  7548

1104   nPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPhePhe.  1120
      :::::  :::::  |||  :::::  :::::  :::::  :::::
      CATTGCTTCTACCATPACAGCTCTTAACAAAAAGGCTAAAGGCGATTGCTTC  7598

7549   CATTGCTTCTACCATPACAGCTCTTAACAAAAAGGCTAAAGGCGATTGCTTC  7598

1121   ...ThrProGlyLysProLysThrAsnValLeuGlyAlaValAsnLys  1136
      :::::  :::::  :::::  |||  :::::  :::::  :::::
      AGTGGAAATACAGCTGAGATACCGGTATATCTTGAACCACTTTGAGCG  7648

7599   AGTGGAAATACAGCTGAGATACCGGTATATCTTGAACCACTTTGAGCG  7648

1137   ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGl  1153
      |||::|  :::::  :::::  :::::  :::::  :::::  :::::
      AAATTGCAAGAGCAAGAAACAACAAGAACTTATGCGCAGATGGA  7698

7649   AAATTGCAAGAGCAAGAAACAACAAGAACTTATGCGCAGATGGA  7698

1153   uThrValSerAsnAlaSerSer.....SerSerAsnProSerSerP  1167
      |  ::  :::::  |||  :::::  :::::  :::::  :::::

```

```

7699 GCACCGCCCTAGTCTAGTGTAAATGGAGGAAGAAATGCTAGACTTT 7744
1167 roGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSer... 1182
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
7749 TAGGCATACTCAAAACTGTTCAAGATGAATCCAAGAACTCCAATCAAG 7798
1183 .....GluAsnGlu..... 1185
    |||||
7799 ATCAAAATGCTAGAAAATGAATGAATCTGAACTTGGTCAAGATGATGCCATGCA 7848
1186 .....AspThrThrMetSer..... 1190
    |||
7849 CAAAGTGAAAAAGTGCATTTTGCACAGACAACTACTTAAGCAGAAACG 7898
1191 .....SerProLeuProGlyLysLysSerAspLysArgAsp 1202
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
7899 CCGAAGCTGAGCTAAACCATGACAGGTGAACACTACTATAAAAGCAAGAT 7948
1203 AspSerAspLeuValArgSerGluLeuGlu.....LysProArgG1 1216
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
7949 AATCTTCAGGCTGCAATGAAAGAAATAGAAAACCTACAGAAAATGTTGTC 7998
1216 yArgLysLysThrProValThrGluGluGluGluLysLeuGlyMetAspA 1233
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
7999 CAAAGTGCAGTACCATATAAGAGAAATTTGACAACTT..... 8038
1233 spLeuThrLysLeuVal.....GlnGluGlnLysProLysGlySerGln 1247
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
8039 ..AAACTAGGTGGTAAAGATTGAAATGGAAATAATAAAGTACTCAAAA 8086
1248 ArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
8087 GCAACACACCAAGAGATTGCCCTACTTAAAGTCTTGTGGAGATAAAGGA 8136
1264 pProGluGluLysArgLeuLysGluAspIle.....LeuGluA 1277
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
8137 AGAAGGCTTGGCTAGGTATAAGAGAACTTAGGCGAGCACAGGCAGACA 8186
1277 snGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPro 1293
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
8187 ACGATACACAGATTGTGTTGCCAAAGATATTACAAAGAGCTTCAACTTTC 8236
1294 ProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysTh 1310
    |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
8237 CCTGTGACTTGTGTTGGTGAAGCGGTATGTCAGAGCACAGCAATGCT 8286
1310 rSerLysLysGlySerLysLysLysSerGlyProProAlaProGluGluG 1327
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
8287 TGTGCTCAG.....TCGGA 8303
1327 luGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSer 1343
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
8304 AAGCGCTTGAAGAGGAGTGCACAT.....TATAGAAG 8341
1344 LysGlnHisArgValSerArgAlaGlnGlnArgAlaGluSerProG1 1360
    |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
8342 AATATCATCATTTATCAGGCATGTCAGATTCTGAAGATCGAAGAA 8391
1360 uSerSerAlaIleGluSerThrGlnSer.....ThrProGlnLysGlyA 1375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
8392 AACAAAGGCAAAATCTGATGTCATTCTCTCATCTGGATCATCACACA 8441
1375 rgGlyArgProSerLysThrProSer.....ProSerGlnPro 1387
    |||||  |||  |||||  |||  |||||  |||||  |||||  |||||
8442 GAGGCTCGCCTCAAAACTTACAGACATGGCCCTTACTTCCA 8491
1388 LysLysAsn 1390
8492 GAAAGTCT 8500
_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT74666
documentation block:

```

```
seq_name: /SIDS1/gcgcdata/geneseq/geneseqn/NA1997.DAT:AAT74666
seq_documentation_block:
```


1833 GAAGAAACACCAGTCCTCCAGTGTAGAACAGGTGGAATTGGTTGGTTTACATC 1882
 423 erAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLys 439
 1883 CAGAAAGCTTTGGTCTCTGTCAGAGAGTTACAGAGGAGAACTCAAG 1932
 440 LeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValG1 456
 1933 CTAAGGAAATCCGGNAGAGAAAGATTGCAAAAGACGCCCTGATTGA 1982
 456 uArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrGluA 473
 1983 GAAG.....TTAGAAATCATGGA 2002
 473 rgMet.....LysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsn 487
 2003 AGCTTGTTCAGAAAACCTTCTCTCGAGAACTCCATATCAGATTGCAAT 2052
 488 Ala.....ValLysAlaLeuAsnGluMetTr 496
 2053 CCTGAGCTGGAACATCAGAGGAAAGCTGAAACATTTGGAGAA..... 2097
 496 pLysCysGlnAsnLeuLeuArg.....HisGlnValLysA 508
 2098 .GCTTCTCTCACTTCTGCTCAAGAGAAATCAGGCCTTCATTCTGAGAAG 2146
 508 sPLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAla 524
 2147 ACATGTTGATTCCCGTCTCAGAGTGCTACAGAAACAGC...AAGAA 2193
 525 IlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLy 541
 2194 CTCTCTGAGAAACATGGTACTGGAGAAATCTCTTTTAATGCCAATGT 2243
 541 sAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAsp..... 555
 2244 AGAACTTGAAGAACTCAAGTCGAAATTTGAAAGCCTAGAGAAATCATGCC 2293
 556AspGluLys..... 558
 2294 ACCTGTGAATGATGACAAGACCACTTTAACTAGTGAAGAGAAAGCCTA 2343
 559IleArgLysGlnLeuGluValLeuValSerPr 569
 2344 CTCTCTCATATTGATCAATGAGAAAGAAATGAAGATCTGGAGAAAGA 2393
 569 oThrCysSerCysLys.....GlnAlaGluG 578
 2394 ACATGCAGAAATTAAGTGAAGATTTTGGAAATTAGCGACTGAGAGGGAGT 2443
 578 lYcysValArgGluIleThrLysLysLeuGly.....AsnProLys 591
 2444 CTTCTCTTCAGAAATTT...GAAGAGTTGGGAGTGTCTTTAAATGCCAAG 2490
 592 GlnProThrAsnProPheLeuLeuMetIleLysPheLeuLeuGluArgI1 608
 2491 GAC.....TGTGAGTAGTGTAGTTTGTGCAATTTTCAGAGATCGGAT 2534
 608 e.....AlaProValHisIleAspThrGluSerIleSerAlaL 621
 2535 GAATGGCATGGAATCAACGATCCATCCACTTCAAGATGAAATCAATGCC 2584
 621 eutLeLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAsp 637
 2585 GGGTGAGAGAAATCAAGTGGAGTGCAGTGCAGTCATGATGCTCACATT 2634
 638 GluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLy 654
 2635 GAGATTATTGTTCTGCAAAATGTTCCAGGACTGGCTT..... 2673
 654 sValLeuSerPheThrHisProLysPheHisSerAlaGluThrPheG 671
 2674GAGAAGAGATT 2683

671 luSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAla 687
 2684 CTTCTTTGATTGCT.....GAGAACTCAGGATTTAAAGAGGCT 2721
 688 AlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPhePr 704
 2722 TCT...AACTGTTAGAGAGCTAGTTTCTGAACTAGAAGAGAA..... 2763
 704 oHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLysG 721
 2764AATATTG 2770
 721 lyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
 2771 GGAACAAGTCAGATTGACTCTCGATTAACTGCAATAAAATACTG... 2817
 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeu..... 751
 2818AGACAGGGATCTACCAGGTATTGATGAAGCTTGAATAAT 2858
 752HisLysSerLeuAspProSerAsn.... 759
 2859 TCCAGGTATAGTTCCGGTGTGATGAGAACTCCGAGACCAGAGAAATATGC 2908
 760LeuGluHisLeuIleThrProLeuValThrIle 770
 2909 ATGATATATTGAATAGACTTGAAGACATGCAAACTATGCTTTTGAGCATT 2958
 771GlyHisIleAlaLeu.....LeuAlaProAspGlnPh 781
 2959 CGGGATCAAAACAGCACTCTGCCATAGAACCTCGTCTCTTATTGAATT 3008
 781 eAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuL 798
 3009 TCTTCGGCAACTTAAATCAGAAGCTGTAGTATTGAAACAGAGAGAGAAA 3058
 798 euMetAsnAspArgLeuPro..... 804
 3059 TCCTTGAGGAAGAACTTGAATCTCAGTGTGAGCAGCTCGTTCGTCGGG 3108
 805GlyLysLysThrThrLysLe 811
 3109 GATGAGACCAGAAACTTATCTTCGTCAACGGAGAAATTGACCACAAAGT 3158
 811 uTrpValProAspGluGluValSerProGluThrMetValLysIleGlnA 828
 3159 T.....AATCAAGGAGTCAACAGAGAAAGGTTTGTGATGTGGAA. 3198
 828 lalleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSer 844
 3199 ..ATTGAGGATTTTCACAGCAGGTGTGAGTTACGATGATGATAC... 3243
 845 LysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAs 861
 3244ACCATTTCTACAGGAGGA 3260
 861 pGlyAsp.....LeuThrGluGlnGlyLysIleSerLysProAspMetS 876
 3261 TAATAACAAAACGTGTGATGAGAAAGCATACTTGACAAAG..... 3300
 876 erArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGlu 892
 3301TCAACATTGCGAGTTGGAGAGGAG 3324
 893 ProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLe 909
 3325 AATGTAACACTAGAA.....GATGATATCTCTTT 3353
 909 uAlaIleAsnAspGluCysTyrGln.....ValArgG 920
 3354 GCTTCTGTCGGAGACTATATATCAAGCAATCTCATCTTCTGTTGGAG 3403

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142923.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 23-APR-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.


```

2449 .....GAGGTAAAGTGTCTC 2465
1000 aHisProAspTyrValLysValGlnAspIleGluLeuLysAspV 1017
      : ||||| : : : : : ||| : : : :
2466 TTGTGATCCA.....CTGGACCTGCTTCAATA 2494
1017 aLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsnGlu 1033
      : : : : : : : : : : : : : : : : : : :
2495 TCAGAACCAACATTCCTGGTATGTTATCTATGTTGCATAGCCTCAACAAT 2544
1034 AsnAsnSerHisAlaPheIle.....ArgLysMetValGluAs 1046
      : : : : : : : : : : : : : : : : : : :
2545 TCTCTCTCAACTGTTCTATGGAGATATATGAGAGGAGATGGTCTGCAA 2594
1046 nLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnG 1063
      : : : : : : : : : : : : : : : : : : :
2595 CCGTATAAACTCAAGAC.....AAGC 2617
1063 LuLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLys 1079
      : : : : : : : : : : : : : : : : : : :
2618 ACAAGATTAAAGAACCTTCCTGGAGGAGCTTAAGAAGGTCTCCACCTCTGTT 2667
1080 SerThrThrTyrSerLeuGlu.....SerProLysAs 1090
      : : : : : : : : : : : : : : : : : : :
2668 AAAACTCTCTCTCCAGCTCCATGCACCACTGTCTCTCACACCTCAAGA 2717
1090 p...ProValLeuProAlaArgPheThrGlnProAspLysAsnPhes 1106
      : : : : : : : : : : : : : : : : : : :
2718 CAAGCCGAAGTGAGCGCTAAGCAGACACCTCTCACTCGACCAACCTTA 2767
1106 exAsnThrLysAsnTyrLeuProGluMetLysSerPheThrPro 1122
      : : : : : : : : : : : : : : : : : : :
2768 ATAACCTCACTGAAGCTG.....TTTGAGAACATCTGCTCTC 2805
1123 GlyLysProLys...ThrThr.....AsnValLeuGlyAlaValAsnLysP 1137
      : : : : : : : : : : : : : : : : : : :
2806 GCAGAATAAGAACATCTACCAAGAACTATTGGACAGA...AATCGTC 2852
1137 roLeuSerSerAlaGlyLysGlnSerGlnThrLysSerArgMetGlu 1153
      : : : : : : : : : : : : : : : : : : :
2853 CGAGAACTTCTACGAGAAGATACTGAAAGACACGACACATCTTATTAAGC 2902
1154 ThrValSerAsnAlaSerSerSerSerSerProGlyArgIle 1170
      : : : : : : : : : : : : : : : : : : :
2903 AGAGCTTCACTAACTTCGTGGAATCTAAGCCGATGATATCAACTCTCTT 2952
1170 eLysGlyArgLeuAsp...SerSer...GluMetAspHisSerGluAsnGlu 1185
      : : : : : : : : : : : : : : : : : : :
2953 AACGATGAATCTAAACGTAAGAAGCTGGAAGAGGACATCAATAAGCTGAA 3002
1186 AspTyrThrMetSerSerProLeuPro..... 1194
      : : : : : : : : : : : : : : : : : : :
3003 GAAGACACTGCAACTGAGCTTCGACCTGTACAAACAGTACAAACTGAAC 3052
1195 .GlyLysLysSerAspLysArgAspAspSerAspLeuVal.....A 1208
      : : : : : : : : : : : : : : : : : : :
3053 TCGAGAGACTCTTCGACAGAGAAGACAGTTCGGCAAGTATTAAGTAGCAG 3102
1208 rgSerGluLeuGluLysProArgGlyArg.....LysLysThrProVal 1222
      : : : : : : : : : : : : : : : : : : :
3103 ATCAAGAAGTTCACTGTCTGCTCAAGGAGCAGCTTGAAAGCAAACTCAACTC 3152
1223 ThrGluGlnGluGluLysLeuGlyMetAspLeuThrLysLeuValGlu 1239
      : : : : : : : : : : : : : : : : : : :
3153 ACTGAACATCCGAAACACGCTACTCGACAACCTCTCAGTGTGTTCTTCAACA 3202
1239 nGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrA 1256
      : : : : : : : : : : : : : : : : : : :
3203 AGAAGAA.....GGAGCCGAGATCGCCGAGACAGAACACTCTG 3243
1256 laSerGluSerAspGluGlnGlnTyrProGluGluLysArg..... 1269
      : : : : : : : : : : : : : : : : : : :
3244 GAGAACACCAAGATTTCTTCAACACATACAAAGGCGCTCGTCAAGTATTA 3293

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT: AAC45028

seq_documentation_block:

ID_AAC45028 standard; DNA; 5414 BP.

XX AAC45028;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 45029.

XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.


```

155870 GGGGCTTACAATTCCTT..... 155886
863 AspleuThiGluGlnGlyLysIleSerLysProAspMetSerArgLeuAr 879
155887 GATCTGATTCCAGGAAGCAATC..... 155910
879 gLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrH 896
155911 .....GGTTTGATGAAGCGGTGATAGTTTCAATACC 155944
896 isGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsn 912
155945 GCAGAGGCTATAAATCTCCACCTAC..... 155970
913 AspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysG1 929
155971 ...GCAACCTGGTGGATCCCGCAGCAATACAGCTCGATTCCGATCA 156017
929 yLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuC 946
156018 GGCCTGTACCATCCGCAATCCGTA..... 156042
946 ysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
156043 .....CATATGATTGAACCATC 156060
963 ValLysAsnIleAsnValArgGluTyrLeuLysGln.....HisAl 977
156061 ACAAAGATGAACCGCATCTCGCCCAACACCTTCAAGAAACCGCGAAGA 156110
977 aAlaValSerGluLysLeuSerLeu.....LeuProGluTyrValV 992
156111 ACCCGATTCCGCNAATTCGCCAACTGATGCAGATGCCCGAAGACAAAA 156160
992 al.....ProTyrThrIleHisLeuLeu 999
156161 TCCGCAAAATCATGAAAATCGCCAAAGAGCGGATTCGATGGAACCCCC 156210
1000 AlaHisAspProAspTyrValLysValGlnAspIle..... 1011
156211 ATCGGCGACGACGACGATTCGCATCTGGGCGACTTCATCGAAGATGCCAA 156260
1012 .....GluGlnL 1014
156261 CAATGTTGGCGGCCGATCGGCAATGTACACCGCTCGACGAAGTAA 156310
1014 euLysAspValLysGluCysLeuTyr.....PheValLeuGlu 1026
156311 CCAAGAAATCTCGAAAGCTGCACCCGCTGAGGCAAAAGTCTCGCT 156360
1027 IleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPhe..... 1039
156361 ATCGGTTTCGGCATCATATGAACCGCACACGCTGGAAGAAGTCGG 156410
1040 .....IleArgLysMetValGluAsnIle.....LysG 1049
156411 CAGACATTTGACGTATACGCGGACGACCATCCGCAAAATCGAGCAAAAG 156460
1049 lnThrLysAspAlaGlnGlyPro..... 1056
156461 CACTCCCAAGCTCGCGCATCCGACAAAGAGCGACCGTTTGAGAAGTTTC 156510
1057 .....AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAs 1070
156511 TTGGACAGCGAAGACAGCAAGCTGTAAACCAAAACCGCAGGTTTCAAA 156560
1070 pValAlaMetAsnIleIleMetSerLysSerThrTyr..... 1083
156561 TACCTGGGTTTTTCTTACACATAAACACGCTTCCACATATCCCA 156610
1084 .....SerLeuGluSerProLysAspProValLeuPro 1094

```

```

156611 CTCCTATCCGAGACCTTTCAAAAATCCCAAAATCCCTTAAATCCCA 156660
1095 AlaArgPhe.....PheThrGlnProAs 1102
156661 CCAAGACATTTAGGGGATTTTCCATGAGCACCTTCTTTTCAGCAAAACGCA 156710
1102 pLys.....AsnPheSerAsnThrLysAsnTyr..... 1111
156711 CAAGCCATGATTGCCAAACACATCGACCGTTCCTCCACTATTGAAGTTGGA 156760
1112 .....LeuProGluMetLysSerPheThrProGlyLys 1124
156761 TCAGTAATTTGTTGGCAACCGATCGAACAGTACCTGAACCGTCAAGAA 156810
1125 ProLysThrThrAsnValLeuGlyAlaValAsnLys...ProLeuSerSe 1140
156811 CCGCTTACCTTCGAGACACCGCGCGCTCCCGCTATCCCGCTCTCTCC 156860
1140 rAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerA 1157
156861 ATGTTCAAAGCCCTCTCTCGGACAATGGACAGCCTCTCCGATCCCGA 156910
1157 snAlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArg 1173
156911 ACTCGAACACAGCCTCATCCCGCATCGATTTCAACCTGTTTTCGCGTT 156960
1174 Leu...AspSerSerGluMetAspHisSerGlu..... 1183
156961 TTGACGAACCTGAGCATCCCGGATTACAGCACCTTATCCCGCTACCGCAAC 157010
1184 .....AsnGluAspTyrThrMetSerSerProLeu..... 1193
157011 TGGCTGCGCGACAGACACACCTGTCCGAACCTGTGGAACTGATTAACTG 157060
1194 .....ProGlyLysLysSerAspLysArgAspAspSerAspL 1206
157061 CCAACTGACCGAAAAAGGCTTAAAGTAGAGAAAGCA...TCCGCGCCG 157107
1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
157108 TCGTTGATGCCACCATTTATTCAGACCGCTGCGACAAACACGCTCAGGCC 157157
1223 ThrGluGlnGluLysLeuGlyMetAspAspLeuThr...LysLeuValG 1239
157158 ATAGAAGTCGATGAAGAGGACAGTACGCGGCCAAACACACACCGAGTAA 157207
1239 lnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThr 1255
157208 GGACAGCGATGCCGCTTGA.....TCAAGAAAAACGGCCTCTACA 157248
1256 AlaSerGluSerAspGluGlnGlnTyrProGluLysArgLeuLysG1 1272
157249 AACTCGGTTACAAACACATACCGCTACCGATCGGAAGCTATATCAG 157298
1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyL 1289
157299 AAACCTGC.....ACATTACCC..... 157314
1289 ysArgGlyArgProProLysProLeuGlyGlyThr..... 1301
157315 .....CGCCCAATGCCATGATGCAAAACACCTGTGCGCGTTGT 157353
1302 .....ProLysGluGluProThrMetLysThrSerLysG1 1314
157354 TGAAGGTTTACCGGAAGTAGCAGCGCTCTATGCCGACA...AAGCTAT 157400
1314 ySerLysLysSerGly 1320
157401 GACAGTCGGAAAAACCGGC 157419
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV21451
seq_documentation_block:

```



```

91490 CAATGTTGCGCGCGGATGTCGGCAATGTACACCGCTGTACAGGTA 91539
1014 euLysAspValLysGluCysLeuTrp.....PheValLeuGlu 1026
|||:||||:||||:||||:||||:||||:||||:||||:||||:
91540 CCAAGAAATCTCGAAGCCTGACACCGCTGAGCGAAAGTCTCGGT 91589
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1027 lleLeuMetAlaLysAsnGluAsnSerHisAlaPhe..... 1039
|||:||||:||||:||||:||||:||||:||||:||||:||||:
91590 ATGCGTTTCGGCATGATGAACACCGACACACGCTGGAGAAGTCGG 91639
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1040 .....lleArgLysMetValGluAsnIle.....LysG 1049
||||:||||:||||:||||:||||:||||:||||:||||:||||:
91640 CAGACAGTTTGACGTAACCGCGCGACGATCGACAAATCGAGGCAAA 91689
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1049 InThrLysAspAlaGlnGlyPro..... 1056
|||:||||:||||:||||:||||:||||:||||:||||:||||:
91690 CACTCGCGAAGTGGCGCATCGGACAAAGCGACCGTTTGAGAAGTTC 91739
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1057 .....AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAs 1070
||||:||||:||||:||||:||||:||||:||||:||||:||||:
91740 TTGGACAGCAAGACAGCAAGCTGTAACCAAAACCAACCGCAGGTTTCAA 91789
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1070 pValAlaMetAsnIleIleMetSerLysSerThrThrTyr..... 1083
||||:||||:||||:||||:||||:||||:||||:||||:||||:
91790 TACCTGCGGTTTTTCTTACACAATAAACACGCTTCCACATATCCCA 91839
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1084 .....SerLeuGluSerProLysAspProValLeuPro 1094
||||:||||:||||:||||:||||:||||:||||:||||:||||:
91840 CTCCTATCCCGAGACTTGTGCAAAATTCCTCAAAATCCCTAAATCCCA 91889
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1095 AlaArgPhe.....PheThrGlnProAs 1102
|||:||||:||||:||||:||||:||||:||||:||||:||||:
91890 CCAAGACATTTAGGGATTTTCCATGAGCACCCTTCTTCAGCAACCGCA 91939
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1102 phys.....AsnPreSerAsnThrLysAsnTyr..... 1111
|||:||||:||||:||||:||||:||||:||||:||||:||||:
91940 CAAGCCATGATGGCCAAACATACACATCCCGTTTCCCACTATTGAATGGA 91989
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1112 .....LeuProProGluMetLysSerPhePheThrProGlyLys 1124
|||:||||:||||:||||:||||:||||:||||:||||:||||:
91990 TCAGTAATTGATGGCAACCGCATCGAACAGTACCTGAACCGTCAAGAA 92039
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1125 ProLysThrThrAsnValLeuGlyAlaValAsnLys....ProLeuSerSe 1140
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92040 CCGGTTACCTTCGAGACCAACCGCGCGCTCCCGCTATCCCTGTGTCC 92089
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1140 rAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerA 1157
||||:||||:||||:||||:||||:||||:||||:||||:||||:
92090 ATGTTCAAGCGCTCTGCTCGGACAATGGCACAGCCTCTCCGATCCCGA 92139
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1157 snAlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArg 1173
||||:||||:||||:||||:||||:||||:||||:||||:||||:
92140 ACTCGAACACAGCCTCATACCCCGCATGATTTCACCTGTTTTCGGT 92189
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1174 Leu...AspSerSerGluMetAspHisSerGlu..... 1183
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92190 TTGACGAAGTACGATCCCGATTCACAGCACCCTTATGCGCTACCGCAAC 92239
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1184 .....AsnGluAspTyrThrMetSerSerProLeu..... 1193
||||:||||:||||:||||:||||:||||:||||:||||:||||:
92240 TGGTGGCGCAAGACGACACCTGTCGCAACTGTTGGAACTGATTAAGTG 92289
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1194 .....ProGlyLysLysSerAspLysArgAspSerAspL 1206
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92290 CCAACTGACCGAAAAGGCTTAAAGTAGAAGACGA...TCCGCGCGCG 92336
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
||||:||||:||||:||||:||||:||||:||||:||||:||||:
92337 TCGTTGATGCCACCATTTATTCAGACCGCTGGCAGCAACAGCGCTCAGGCC 92386
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1223 ThrGluGlnGluLysLeuGlyMetAspLeuThr...LysLeuValG 1239
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92387 ATAGAAGTCGTGAAGAGGACAAAGTCAGCGGCCAACACACCGAGTAA 92436

```

```

1239 InGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThr 1255
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92437 GGACAGCGATGCCGTTGGA.....TCAGAAAAACGGCCTCTACA 92477
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1256 AlaSerGluSerAspGlnGlnTrpProGluGluLysArgLeuLysG 1272
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92478 AACTCGGTTACAACAACATACCCGTACCGATCGGAGGCTATATCGAG 92527
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyL 1289
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92528 AAATGTC.....ACATTACCC..... 92543
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1289 ysArgLysArgProProLysProLeuGlyGlyThr..... 1301
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92544 .....CCGCCATGCCCATGAGTGCAACACACCTGTGCGCGTTGT 92582
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1302 .....ProLysGluGluProThrMetLysThrSerLysLysG 1314
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92583 TGAAGGGTTACCGAAGGTACGACCGTCTATGCGGACA...AAGGCTAT 92629
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1314 ysSerLysLysSerGly 1320
|||:||||:||||:||||:||||:||||:||||:||||:||||:
92630 GACAGTCGGAACCCGGC 92648
|||:||||:||||:||||:||||:||||:||||:||||:||||:

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA81489

seq_documentation_block:

ID: AAA81489 standard; DNA: 837096 BP.

XX AAA81489;

DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;

XX WPI: 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -

PS Claim 7; Page 629-865; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF

296 uAlaLysMetPheGlyAlaLysAspSerGluLeuAla.....Serg 310
I : : : : :
89114 TAGCGAGCTGGTCGCAGCATCGACCCGCCACAACCTCGCGCAACTGCTAGGAC 89163

310 InAsnLysProLeuTrpGlnCysTyrLeuGlyAtgPheAsn..... 323
I : : : : :
89164 AGGAAGCCCGCAAGCCGCATCGTCAAAACAAAAAACATCAAACTGCCTCCG 89213

324 ...AspIleHisValProIleArgLeuGluCysValLysPheAlaSerHi 339
: : : : :
89214 ATTTCGGTCAAACAGCCCGCTCATGTGTCAGCTGGTGATCAGCGCAAAATCCG 89263

339 sCysLeuMetAsnHisProAspLeuAlaLys.....AspLeuThrGluT 354
C : : : : :
89264 CAGCTCTTTGATAAATCCGGAATTGGCGTCATATATAGCATGCTGCCCGATT 89313

354 yrLeu.....Lys 356
I : : : : :
89314 ATTCGGCGTTGGACGGTGATTTCGCCCTGCCATCTTCCCAATCTTCCCAATCG 89363

357 ValArgSerHisAsp.....ProGluGluAla..... 365
: : : : :
89364 ATTAAAACCACATGCGCGCTACCGAGAACCCTCAGGTTTTAGAGTATAT 89413

366IleArgHisAspValI 371
I : : : : :
89414 GCGCGGCTCGCTTAGGAAGAAACGATNACCCGGAATCTTCCTATTCAACGC 89463

371 leValSer.....IleValThrAlaLysLysAspIle..... 382
I : : : : :
89464 ACCAATCGGAAGAAATGAACAGCAGCAGTGAAGAAGATTGCGAGAATTC 89513

383LeuLeuValAsnAspHisLeuLeuAsnPheValAr 394
I : : : : :
89514 CAATTCGCATGAAACAACTGCTCAATGAG....TTAAATACAGCCA 89557

394 gGUARgthrLeuAspLysArgTrrAgvalargLysGluAlaMetMetG 411
I : : : : :
89558 AATCGAAACATTAAAA.....CAAAAAGCTGCATCCG 89592

411 lyLeuAlaGlnIleTyrlLysLysTyrAlaLeuGlnSerAlaAlaGlyLys 427
I : : : : :
89593 GCTTAATGAAAGCGAGAAACAACTTTGCTGCTGCTGCACCGCAAA 89642

428AspAlaAlaLysGlnIleAlatrpIleLysAspLy 439
I : : : : :
89643 CAAATTCACCGCGGATTCGCCCATCGTAAACCGTTATCCGCTCGAA 89692

439 sLeuLeuHisIleTyrlTyrGlnAsnSerIleAspArgLeuLeuValG 456
I : : : : :
89693 AAGCATTCACCGCGCTGCAACACAGCACACC..... 89723

456 luArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGlu 472
: : : : :
89724TGCAGAACCCATCCCCAAAGCGCTTCAGACGGCATCA 89762

473 ArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaVa 489
I : : : : :
89763 GAGTACCTACTCTGCCAGCGCTTCAGGTGGCTCCAAACGCAACCGTCG 89812

489 lLysAlaLeuAsnGlu.....MetTrpLysCysGlnAsnLeuLeuA 503
I : : : : :
89813 GCATCTTACCAACAGAAAGCAGACAATGTCCAGAAACCAAAATCAGCAAG 89862

503 rgHisGln..... 505
I : : : : :
89863 ANTATCAGACGACACCGCTCGCTTAAAGCATTTGAAGACCAACGCGCGGC 89912

505 505
CTGGCTCAGCTCATCATCGGTAAAGAACCGCGGTACATCACTACTC 89962


```

111 ||| ||| : : : : : |||
7916 CCGGTTACCTTCAGACACCGCGCGCTCCCGCTATCCCTGCTGTC 7965
1140 rAlaGlyLysGlnSerGlnThrLysSerArgMetGluThrValSerA 1157
: : : : : ||| : : : : : |||
7966 ATGTTCAAAAGCGTCTGCTCGGACAAATGCACAGCTCTCCGACCGA 8015
1157 snAlaSerSerSerAsnProSerProGlyArgIleLysGlyArg 1173
||: : : : : ||: : : : : ||: : : : : ||: : : : :
8016 ACTCGAACACAGCTCATCCCGCATCGATTTCACCTGTTTGCCTT 8065
1174 Leu.. AspSerSerGluMetAspHisSerGlu..... 1183
||| : : : ||| : : : : : |||
8066 TTGACGAACATGACATCCCGATTACAGCACCTTATGCCGTACCGCAAC 8115
1184 ..... AsnGluAspTyrThrMetSerSerProLeu..... 1193
8116 TGCGTGGCGCAAGACACACCTCTGCCGAACCTGTGGAACATGATTAAC 8165
1194 ..... ProGlyLysLysSerAspLysArgAspSerAspL 1206
||| ||| ||| : : : |||
8166 CCAACTGACCGAAAGGCTTAAGTAGAAGCA... TCGCGCGCG 8212
1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
: : : : : ||| : : : : : |||
8213 TCGTTGATGCCACCATATTACAGCCGCTGGCAGCAACAGCGCTCAGGCC 8262
1223 ThrGluGlnGluLysLeuGlyMetAspAspLeuThr.. LysLeuValG 1239
||| : : : : : ||| : : : : : |||
8263 ATAGAAGTCGTGAAGAAAGGACAAAGCAGCGGCCAAACACACCGAGTAA 8312
1239 lngGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThr 1255
: : ||| ||| : : : |||
8313 GGACAGCGATGCCCTTGA..... TCAGAAAAACGGCCTTACA 8353
1256 AlaSerGluSerAspGluGlnIntrProGluLysArgLeuLysG 1272
||| : : : : : ||| : : : : : |||
8354 AACTCGGTTACAAACAACATACCCGTACCGTACCGTACCGAAGGCTATACGAG 8403
1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyL 1289
: : : : : |||
8404 AAATGC..... ACATTACCC..... 8419
1289 ysArgGlyArgProProLysProLysProLysGlyGlyThr..... 1301
8420 ..... CCGCAATGCCATGATGATGCAACACCTCTCCCGTTGT 8458
1302 ..... ProLysGluGluProThrMetLysThrSerLysLysG 1314
||| ||| : : : : : ||| |||
8459 TGAAGGGTTACCGAAGGTACGACCGCTCTATCCGACA... AAGGCTAT 8505
1314 ySerLysLysLysSerGly 1320
: : : : : |||
8506 GACAGTCGGAAACCGGC 8524

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAF21611

seq_documentation_block:

ID AAF21611 standard; DNA; 349980 BP.

XX AC AAF21611;

XX DT 13-MAR-2001 (first entry)

XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.

XX OS Neisseria meningitidis.

XX PN WO20006791-A1.

```

XX 09-NOV-2000.
PD
XX 08-MAR-2000; 2000WO-US05928.
PF
XX 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
DR
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX Claim 7; Appendix A; 692pp; English.
XX

```

The present invention describes the full length genome of *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613 represent fragments of the NMB genomic sequence, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. *Neisseria* nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to *Neisseria* bacteria or as a diagnostic reagent for detecting the presence of *Neisseria* bacteria or of antibodies raised to *Neisseria* bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines than the outer membrane proteins currently used.

XX SQ Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 195.50 | Length: | 1676 |
| Ratio: | 0.273 | Gaps: | 81 |
| Percent Similarity: | 42.780 | Percent Identity: | 18.795 |

alignment_block:

US-09-512-581-2 x AAF21611 ..

Align seg 1/1 to: AAF21611 from: 1 to: 349980

```

13 IleThrTyrProGlyValLysGluIleSerAspLysIleSerLysG 29
: : : : : ||| : : : : : |||
88107 CTGCTGTTCCGGAGCGGTTCAGTTCCTTCCGACCGCGTGGGT.... 88151

29 uGluMetValArgLeuLysMetValValLysThrPheMetAspMetA 46
: : : : : |||
88152 .....ATGGTCGTGCGCAAGAGTCACGGGCAAA 88179

46 spGlnAspSerGlu.....GluGluLysGluLeuTyrLeu 57
: : : : : |||
88180 ACCATAATCCGAAGTCGCGCAAGCTAAGAAAAACACGACACTG 88229

58 AsnLeuAlaLeuHisLeuAlaSerAspPheLeuLys..... 70
: : : : : |||

```


7870 CTTCTGTAGAGAAGTCGACTGACAGGCCAGAGAGGGCCTGTAGGGTCAA 7919

1289 sArgGlyArg...ProProLysProLeuGlyGly.GlyThrProLysGlu 1304

7920 GCGAGGCCGCACTGTCCCCAGCCCCAAGGTGGACTGGAGTCCAAAGGCCA 7969

1305 GluProThrMetLysThrSerLys 1312
 |||||:::~::~:
7970 GTGAGAACTGTAAAGTCCAGTGAA 7993

seq_name: /SIDS1/qcdata/geneseq/geneseq/NA2000.DAT:AAA81482

seq_documentation_block:

| ID | AAA81482 standard; DNA; 14652 BP. |
|-----|-----------------------------------|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 8 |
| 9 | 9 |
| 10 | 10 |
| 11 | 11 |
| 12 | 12 |
| 13 | 13 |
| 14 | 14 |
| 15 | 15 |
| 16 | 16 |
| 17 | 17 |
| 18 | 18 |
| 19 | 19 |
| 20 | 20 |
| 21 | 21 |
| 22 | 22 |
| 23 | 23 |
| 24 | 24 |
| 25 | 25 |
| 26 | 26 |
| 27 | 27 |
| 28 | 28 |
| 29 | 29 |
| 30 | 30 |
| 31 | 31 |
| 32 | 32 |
| 33 | 33 |
| 34 | 34 |
| 35 | 35 |
| 36 | 36 |
| 37 | 37 |
| 38 | 38 |
| 39 | 39 |
| 40 | 40 |
| 41 | 41 |
| 42 | 42 |
| 43 | 43 |
| 44 | 44 |
| 45 | 45 |
| 46 | 46 |
| 47 | 47 |
| 48 | 48 |
| 49 | 49 |
| 50 | 50 |
| 51 | 51 |
| 52 | 52 |
| 53 | 53 |
| 54 | 54 |
| 55 | 55 |
| 56 | 56 |
| 57 | 57 |
| 58 | 58 |
| 59 | 59 |
| 60 | 60 |
| 61 | 61 |
| 62 | 62 |
| 63 | 63 |
| 64 | 64 |
| 65 | 65 |
| 66 | 66 |
| 67 | 67 |
| 68 | 68 |
| 69 | 69 |
| 70 | 70 |
| 71 | 71 |
| 72 | 72 |
| 73 | 73 |
| 74 | 74 |
| 75 | 75 |
| 76 | 76 |
| 77 | 77 |
| 78 | 78 |
| 79 | 79 |
| 80 | 80 |
| 81 | 81 |
| 82 | 82 |
| 83 | 83 |
| 84 | 84 |
| 85 | 85 |
| 86 | 86 |
| 87 | 87 |
| 88 | 88 |
| 89 | 89 |
| 90 | 90 |
| 91 | 91 |
| 92 | 92 |
| 93 | 93 |
| 94 | 94 |
| 95 | 95 |
| 96 | 96 |
| 97 | 97 |
| 98 | 98 |
| 99 | 99 |
| 100 | 100 |

AC AAA81482:

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence qnm_30 SEQ ID NO:30.

Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
antigen; vaccine; diagnosis; infection; antibacterial; identification;
Meningococcus B; MenB; ds.

Neisseria meningitidis.

AA
PN
WO200022430-A2.

XX
PD
20-APR-2000-

XX
PF 08-OCT-1999; 99WO-US23573.

XX
PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.

(CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Rappuoli R, Pizza M;

DR WPI; 2000-318079/27.

Isolated nucleotide sequences of *Neisseria meningitidis* which can be used in the diagnosis and treatment of *N. meningitidis* infection and other *Neisseria* infections, for example, *N.gonorrhoea* -

PS Claim 7; Page 582-586; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AAAB1453 to AAAB2414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding proteins; AAAB1254 to AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAAB1322 to AAAB1452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisserial* bacteria. For example, some of the identified proteins could be components of vaccines against *Neisseria meningitidis*; against all serotypes and/or against all pathogenic *Neisseria*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Neisseria meningitidis* B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 14652 BP; 3745 A; 4117 C; 3436 G; 3354 T; 0 other;

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 195.50 | Length: | 1676 |
| Ratio: | 0.273 | Gaps: | 81 |
| Percent Similarity: | 42.780 | Percent Identity: | 18.795 |

alignment_block:

US-09-512-581-2 x AAA81482

Align seq 1/1 to: AAA81482 from: 1 to: 14652

13 IleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysCl 29
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
3983 CTGTCTGTTCCGGAGCGGGTTCAGTTCCTTGCGACCGGTGGGT 4027

29 uGluMetValArgLeuLysMetValValLysThrPheMetAspMetA 46
 |||||
 4028ATGGTCGTCGCCGAAAGTGCACGGGCAAA 4055

46 spGlnAspSerGlu.....GluGluLysGluLeuTyrLeu 57
 ::::~::~||| ::::~::~|||
 A056 ACCATAATCCCGACACTCCTCCGCCCAACCTACAAAAAAGCCACACACATC 4105

[illegible]

```

71 .HisProGlyLys.....AspValArgLeuLeuValAlaCysC 83
   ::::::::      :::::  :::::  :::::  :::::

```

83 ysLeuAlaAspIlePheArgIleTyrAlaProGluAla..... 95
 ::||| |||||:::~::~:

96ProTyrThrSerPro..... 100
||||| :::

101AspLysLeuLysAspIlePheM 108
|||:::|
4250 ACGCCAGAGTGTCTCCATACCGTAATCCTAATAATACGCGTTAGTGGATACGGGGGAI 430

```

108 etPheIleThrArgGlnLeuLysGly..... 116
||||| |||::: ::|||
4306 GGTGATTGACAATGAGGACGGCATTACGACCGGCTTCCGCCCATCGGATTA 4355

```

117 ...LeuGluAsp^{Thr}Lys.....SerProG1 124
 4356 TGT^TTCC^CCA^TCCG^CCA^TCCG^CCGG^CCA^GTTAT^CGG^TTTT^TCG^CGC^GCA^GG 4405

4406 GTGCTGGACGACTCGAAGCCGAAATATTAAATTCTCCCGATACGCCCTTT 4455
124 nPheAsnArg.....TyrPheTyrLeuLeuGluAsnIleAlaTrpV 138

4456 GTTCGATAAGGGGAAAAACCTTTACGGACTGTATGAAGGGCGTGCCGCTG 4505
138 aLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePhe 154

4506 TCAAG.....GAAGCGGGCGGATTTGGTG 4531
155 ThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGly..... 168

4532 GTCGAAGGCTATATGGACGTGGTCGGCGTGGCACAGTTCGGCGTGGGCTA 4581
..... |||
169HisAsnGlnLysValH 174

4582 CGCGGTGGCGGCTTTGGGTACGGCGACGCGGGGAACACAGTCAAAATCC 4631
 174 {sMetHjsMetVa|AsnLeuMetSerSerIleT}eCvs G|uG|vAsn 189

OM of: US-09-512-581-2 to: N_Geneseq_0601.* out_format : pfs

Date: Sep 26, 2001 12:37 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cyn2.1/USPTO.spool/US09512581/runat_25092001_095747.18412/app_query.fasta_1.1470
-DB=N_Geneseq_0601 -OFM=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum2
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09512581 -CGN1_1_330 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-512-581-2

Query length: 1391

Database: N_Geneseq_0601.*

Database sequences: 730101

Database length: 313950809

Search time (sec): 187.440000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|--|-----------|---------|----------|-------|---------------|
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA28051 + | 7193.00 | 8969.67 | 0.0 | 5271 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA28052 + | 7193.00 | 8969.51 | 0.0 | 5355 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA7423 + | 2932.50 | 3643.34 | 5.8e-195 | 3957 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA18338 + | 762.50 | 941.01 | 1.9e-44 | 1217 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV87629 + | 698.00 | 870.72 | 1.6e-40 | 439 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV21451 + | 609.00 | 754.10 | 4.9e-34 | 738 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA242861 + | 513.00 | 641.59 | 9.1e-28 | 351 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA202187 + | 501.00 | 628.35 | 5.0e-27 | 295 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA280598 + | 349.00 | 432.25 | 4.2e-16 | 530 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV14598 + | 225.00 | 253.62 | 3.7e-06 | 5334 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV21511 + | 207.00 | 228.70 | 9.1e-05 | 6755 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AAV34578 + | 201.50 | 217.69 | 0.0004 | 10136 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA13177 + | 197.50 | 220.21 | 0.0003 | 4842 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ6851 + | 197.00 | 213.51 | 0.0006 | 8789 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV09076 + | 195.50 | 213.51 | 0.0006 | 8789 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA181482 + | 195.50 | 206.43 | 0.0016 | 14652 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV21611 + | 195.50 | 174.10 | 0.0959 | 34980 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV21611 + | 195.50 | 165.21 | 0.3120 | 83706 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV21451 + | 193.00 | 214.38 | 0.0006 | 4940 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA45028 + | 193.00 | 213.44 | 0.0006 | 5414 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AA226819 + | 192.00 | 206.33 | 0.0016 | 9626 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV74666 + | 191.00 | 204.37 | 0.0021 | 10320 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ89814 + | 190.50 | 219.43 | 0.0003 | 2212 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ67602 + | 189.50 | 210.15 | 0.0010 | 4868 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ45991 + | 188.50 | 209.28 | 0.0011 | 4689 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:AAQ90918 + | 186.50 | 202.72 | 0.0025 | 6981 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA337108 + | 186.00 | 213.54 | 0.0006 | 2271 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:AAV54420 + | 186.00 | 213.53 | 0.0006 | 2272 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AA13176 + | 186.00 | 205.80 | 0.0017 | 4851 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV18885 + | 185.50 | 194.52 | 0.0073 | 13815 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:AAV28296 + | 185.00 | 211.05 | 0.0009 | 2565 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AAV58751 + | 184.50 | 204.08 | 0.0021 | 4779 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AAV27558 + | 183.50 | 188.61 | 0.0155 | 19307 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ98079 + | 183.00 | 197.67 | 0.0049 | 7458 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV90138 + | 181.50 | 209.52 | 0.0011 | 1939 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:AAV90338 + | 181.50 | 190.19 | 0.0127 | 12923 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV23545 + | 181.00 | 196.41 | 0.0057 | 6605 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV74903 + | 181.00 | 196.41 | 0.0057 | 6605 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV01731 + | 180.00 | 206.79 | 0.0015 | 2108 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ86236 + | 180.00 | 198.96 | 0.0041 | 4548 | |

| | | | | | |
|--|--------|--------|--------|------|--|
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV91039 + | 180.00 | 192.77 | 0.0091 | 8351 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV01274 + | 180.00 | 192.77 | 0.0091 | 8351 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ54841 + | 179.50 | 195.00 | 0.0068 | 6106 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV78867 + | 178.00 | 193.38 | 0.0084 | 6152 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV77783 + | 177.50 | 192.07 | 0.0100 | 6579 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV77782 + | 177.50 | 191.70 | 0.0104 | 6819 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV65853 + | 177.00 | 190.93 | 0.0115 | 6921 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV00768 + | 177.00 | 187.51 | 0.0179 | 9679 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV47245 + | 176.00 | 195.50 | 0.0038 | 2638 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV46186 + | 176.00 | 196.37 | 0.0057 | 3586 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV46187 + | 176.00 | 196.27 | 0.0058 | 3624 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:AAQ58610 + | 176.00 | 193.58 | 0.0082 | 4716 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV24044 + | 175.50 | 191.81 | 0.0103 | 5276 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV02559 + | 174.50 | 191.79 | 0.0103 | 4677 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:AAV37401 + | 174.00 | 188.41 | 0.0159 | 6519 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAV05113 + | 174.00 | 190.94 | 0.0115 | 4783 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV78868 + | 174.00 | 189.78 | 0.0134 | 5361 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV10741 + | 173.50 | 192.73 | 0.0091 | 3771 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ9326 + | 173.50 | 186.77 | 0.0157 | 6773 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:AAV05301 + | 172.50 | 187.17 | 0.0187 | 5760 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV58901 + | 172.50 | 185.88 | 0.0220 | 6536 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV35363 + | 171.50 | 187.48 | 0.0179 | 4940 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV61982 + | 171.00 | 181.92 | 0.0366 | 8022 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV58840 + | 171.00 | 138.30 | 9.84 | 5800 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV58880 + | 170.50 | 181.37 | 0.0393 | 7962 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ6611 + | 170.00 | 220.49 | 0.0003 | 161 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ75165 + | 170.00 | 179.08 | 0.0527 | 9370 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ75164 + | 170.00 | 179.06 | 0.0528 | 9370 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV50687 + | 169.00 | 192.74 | 0.0091 | 2169 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV31959 + | 168.50 | 189.61 | 0.0137 | 2774 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6106 + | 168.00 | 195.01 | 0.0068 | 1536 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV31965 + | 168.00 | 193.70 | 0.0081 | 1746 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV31407 + | 168.00 | 192.36 | 0.0096 | 1992 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV34457 + | 168.00 | 192.36 | 0.0096 | 1992 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV31966 + | 168.00 | 190.12 | 0.0128 | 2480 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV38817 + | 168.00 | 188.11 | 0.0166 | 3023 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV40200 + | 168.00 | 182.61 | 0.0334 | 5175 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV72945 + | 168.00 | 182.35 | 0.0347 | 5321 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV14584 + | 167.50 | 186.83 | 0.0195 | 3222 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV67602 + | 167.50 | 173.71 | 0.1049 | 1168 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV67602 + | 167.00 | 181.14 | 0.0405 | 5498 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV58909 + | 167.00 | 180.80 | 0.0423 | 5380 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV20248 + | 167.00 | 150.12 | 2.16 | 1113 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV20248 + | 167.00 | 128.71 | 33.71 | 9107 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80911 + | 166.50 | 180.74 | 0.0426 | 5181 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV58846 + | 166.50 | 180.19 | 0.0457 | 5471 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV58846 + | 166.50 | 179.00 | 0.0532 | 6145 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:AAV68343 + | 166.00 | 190.36 | 0.0124 | 1896 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV68343 + | 166.00 | 185.34 | 0.0236 | 3104 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV40172 + | 166.00 | 185.33 | 0.0236 | 3106 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV68758 + | 166.00 | 174.72 | 0.0922 | 8799 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAV39681 + | 165.50 | 179.25 | 0.0516 | 5306 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV70095 + | 165.50 | 175.96 | 0.0786 | 7326 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV88575 + | 165.00 | 181.50 | 0.0387 | 4002 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV60603 + | 165.00 | 181.50 | 0.0387 | 4002 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV17568 + | 164.50 | 189.66 | 0.0136 | 1689 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAV85392 + | 164.50 | 172.72 | 0.1191 | 8906 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:AAQ43526 + | 164.50 | 167.93 | 0.2203 | 1425 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ75181 + | 164.50 | 167.93 | 0.2203 | 1425 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV20477 + | 164.50 | 179.66 | 0.0283 | 4237 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV6647 + | 164.00 | 165.28 | 0.3092 | 1738 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV81493 + | 164.00 | 134.70 | 15.63 | 3499 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV21607 + | 163.50 | 176.44 | 0.0740 | 5471 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV58905 + | 163.50 | 171.87 | 0.1329 | 8364 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV51630 + | 163.00 | 188.18 | 0.0164 | 1625 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ86237 + | 163.00 | 177.80 | 0.0621 | 4500 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:AAQ10378 + | 163.00 | 175.11 | 0.0877 | 5858 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAQ00769 + | 163.00 | 170.47 | 0.1590 | 9241 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:AAQ03358 + | 162.50 | 185.96 | 0.0218 | 1899 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:AAQ01314 + | 162.50 | 180.97 | 0.0413 | 3099 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAV58911 + | 162.50 | 175.17 | 0.0870 | 5477 | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:AAQ45339 + | 162.50 | 167.26 | 0.2400 | 1190 | |

[illegible]

7820 CAGGACCATCTCGGAGAACCCACGAGAAATCCGTCCCGACGTAATTAAT 78659
1278 GluAspGluGlnAsn.....SerProProLysLysGly.....Ly 1289
|||||::: :::: |||:::|||||
7870 CTTCTGAGAGNAGTCCGACTGACAGCCCCACAGAGGGCCTGAGGGTCAA 7919
1289 sArgGlyArg...ProProLysProLeuGlyGly.GlyThrProLysGlu 1304
|||||::: |||:::||||| :::: |||::: |||::: |||:::
7920 GCGAGGCGGACTGTGCCCCAGCCCCAACAGCTGGACTGGAGTCCCAAGGSCA 7969
1305 GluProThrMetLysThrSerLys 1312
|||||:::|||||::: |||::: |||::: |||::: |||:::
7970 GTGAGAAGCTGAAGCTCCAGTGAA 7993

seq_name: /SIDS1/qcdata/qeneseq/qeneseq/NA1998.DAT:AAV09076

seq_documentation block:

seq_documentation_block:
ID AAV09076 standard; DNA: 8789 BP.

AAV09076:

28-MAY-1998 (first entry)

DE Mitosin nucleic acid sequence.

Mitosis; phosphoprotein; mitotic cell cycle; antibody; analogue; inhibition; M phase; Antagonist; hyperproliferative cell; cancer; leukaemia; lymphoma; chromosome segregation; ss.

OS Homo sapiens.

| AA | FH | FT | Key | Location/Qualifiers |
|----|----|----|-----|---------------------|
| | | | CDS | 543..7992 |

```
FT      /tag= a
FT      /product= Mitosin
FT      8740..8746
FT      polyA_signal
```

PN US5710022-A.

20-JAN-1998.

24-OCT-1994: 94US-0328254.

24-OCT-1994: 94US-0328254.

| | | |
|----|--------------|---------------|
| PR | 22-OCT-1993; | 93US-0141239. |
| PR | 24 OCT 1994, | 94US 0320234. |

PA (TEXA) UNIV TEXAS SYSTEM.

PI Lee W. Zhu X:

WPI: 1998-109817/10.

DR WFI; 1998-109617/
DR P-PSDB; AAW23996.

AA
PT
PT
PT
PT

New isolated mitotin protein and gene - useful for, e.g. developing products for therapy and diagnosis of hyper-proliferative disorders such as cancers or psoriasis

PS Claim 1; columns 23-40; 43pp: English.

This sequence encodes mitotin, a phosphoprotein necessary for the cell to enter mitosis. The protein's degradation is also necessary for the cell to advance into the next stages of mitosis. The mitotin protein, can be used to control the growth of cells. An anti-mitotin antibody, a mutant or a non-functional analogue of mitotin can inhibit the mitotic cell cycle by preventing the cells from entering the M phase, and over expression of mitotin or its functional equivalent, would inhibit the cycle by preventing cells from leaving the M phase. Antagonists to this protein can be used to control hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease, psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast cancer, sarcomas and other neoplasms, bladder cancer, colon cancer, lung cancer and various leukemias and lymphomas). Reintroduction or supplementation of lost

[illegible]

| | | |
|------|---|------|
| 7025 | GGAAATGAAGTTGTTG..... | 7041 |
| 999 | LeuAlaHisAspProAspTyrValLysValGlnAspIleGluInLeuLys | 1015 |
| 7042 |ATCAATCTCTGTAACAGACTGGA | 7064 |
| 1015 | AspValLysGluCysLeuTppPheValLeuCluIleLeuMetAlaLys | 1032 |
| 7065 | AGGAAAGAGGATACTGCGAAGAAGACTCTCAACTCAAGCTGCAC | 7114 |
| 1032 | snGluAsnAsnSerHisAlaPheIleArg.....LysMet | 1043 |
| 7115 | AGGAGAGCAGAAAAACAGGTACTGTTATGGATACCAAGTCGATGAATTA | 7164 |
| 1044 | ValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAlaLys | 1060 |
| 7165 | ACAACCTCAGATCAAGAAGACTCAAGAGAACTCTTGAAGAAAAACCAAGGA | 7214 |
| 1060 | sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleLeu | 1077 |
| 7215 | GGCAGATGAATACTTGGATAAGTACTGT.....TCCTTGCTTA | 7252 |
| 1077 | etSerLysSerThrThrTyrSerLeuGluSerProLysAsp.....Pro | 1091 |
| 7253 | TAAGCCCATGAA.....AAGTTAGAGAAGCTAAAGAGATGTTAGAG | 7293 |
| 1092 | ValLeuProAlaArgPhePheThrClnProAspLysAsnPheSerAsnTh | 1108 |
| 7294 | ACAAAGTGGCCCATCTGTGTTCACAGCAATCTAAACAAGATCCCGAGG | 7343 |
| 1108 | rLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysP | 1125 |
| 7344 | GTCTCTTGTGATGCTCCAGTGTT.....C | 7369 |
| 1125 | roLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAla | 1141 |
| 7370 | CAGACCATCTCCAATCCCTTCTGTACTGAAAGAGGTTATCATCTGGC | 7419 |
| 1142 | GlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl | 1158 |
| 7420 | CAAAATAAGCTTCAGGCAGAGCAAGATCCAGTGGATATGGAGAA | 7469 |
| 1158 | aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuA | 1175 |
| 7470 | TGTTGGAGGACCAACACTGTCATCCCAAGAGAGCTTTCTTAAAAAAGCA | 7519 |
| 1175 | spSerSerGluMetAsp.....HisSerGluAsnGluAspTyrThr | 1188 |
| 7520 | AGAAAGCAGTCATCAGTGGTATTACCCCTGCAGAGACACAGGAAGGTACT | 7569 |
| 1189 | MetSerSerProLeuProGlyLysLysSerAspLysArgAspSerAs | 1205 |
| 7570 | GAGTTTCAGCCAGAGGGACTCCAGAAGCTTGTAAGAAAGGGTTTGCTGA | 7619 |
| 1205 | pLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrProv | 1222 |
| 7620 | CATCCCAGACAGGAAGACTAGCCCATATATCTCGAAGAACACCACTGG | 7669 |
| 1222 | alThrGluGlnGlu.....GluLysLeuGlyMetAspAsp | 1233 |
| 7670 | CAACTCGACACAGCCCCCGCTGCTGCACAGAACTTAGCGCTATCCCCA | 7719 |
| 1234 | LeuThr.....LysLeuValGlnGluGlnLysPro...LysG | 1245 |
| 7720 | CTGAGTCTCGCAAGAAAAATCTTGCAGAGTCTCTCAACCAACACAGCTGG | 7769 |
| 1245 | ySerGlnArgSerArgLys...ArgGlyHisThrAlaSerGluSerAspG | 1261 |
| 7770 | TGGCAGCAGATCACAAAGGTCAAGTTGCTCAGCGGACCCAGTAGATT | 7819 |
| 1261 | luGlnGlnThrProGluGluLysArgLeuLysGluAspIleLeuGluAsn | 1277 |

4597 ATGAAGGAGAAACACACAGAGCTTGTAGCTCTCATCAAAAGTGAGTGTCTCCA 4646
 182 rSerIle...IleCysGluGlyAspThrValSerGln...GluLeuLeuA 197
 4647 TTGCATTACAGTGGCAGAGCAGAGGTGAAGAAAGACGGAACCTCTCTC 4696
 197 spThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGln 213
 4697 AGACTTTGTCTCTCTGANGTG.....AGTGAGCTCTTAAAGAC 4734
 214 AlaTyrAspLeuAlaLysAlaLeu.....LeuLysArgThrAlaGlnAl 228
 4735 AAAACTCATCTCCAGGAAAGCTCAGAGTTTGGAAAGGAGCTCACAGG 4784
 228 aileGluProTyrIleThrPheAsnGlnValLeuMetLeuGlyL 245
 4785 ACTGTCTTTGACAAATGTGAGCTGGAAACCAAAATGGCAACAGTGAATA 4834
 245 yst.....ThrSerIleSerAspLeuSerGluHisVal 255
 4835 AAGAGAAAGAAATTCCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 4884
 256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSe 272
 4885 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAAGGCCTTGGAGGCCG 4934
 272 rValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluL 288
 4935 ACTGGTGGAAAGAGTGAAGTTCGCATTGAGGCTGAGCTCAACACAGAGG 4984
 289ArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
 4985 AAGTGCATCAGCTCAGAGAGGCGATCAGAAACTGAGAGTTCGCATTGAG 5034
 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnC 317
 5035 GCCGATGAAAGAGAGAGCTGCATCGCAGAGAAC.....TGAAAGA 5078
 317 ystTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCys 333
 5079 ACGGACGGGAGAGT.....ATTCACTTAAGGATAAAGTTGAGAAC 5122
 334 ValLysPheAlaSerHisCys.....Le 341
 5123 TTGAAGGGAATTCAGATGTGCAGAGAAACAGGAGCTAGTATTCTT 5172
 341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysVala 358
 5173 GATG.....CCGAGAATTCCAAAGCAGAA..GTAGAGACTCTAAACAC 5215
 358 rgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIle 374
 5216 AAATAGAGAGATGCCAGAGCCTGAAAGTTTGTGAATTAGACTTGTGTC 5265
 375 ValThrAlaAlaLysLysAspIleLeuLeuValAsn..... 386
 5266 ACGTTAAGGCTGAAAAGAAATCTGCACAAACAAATACAGAAAAACA 5315
 387AspHisLeuLeuAsnPheValArgGluArgT 397
 5316 AGGTCAGTTGTGACAACTAGACAAAGTTACTCTCTTCAATTTAAAGTCTGT 5365
 397 hrLeuAspLysArgTyrP.....ArgValArgLysGluAlaMetMet 410
 5366 TAGAAGAAAGAGGACAGCAGATACAGATCAAGAAAGAAATCTTAAACT 5415
 411 GlyLeuAlaGlnIleTyrLys.....LysTyrAlaLeuGlnSerAlaAl 425
 5416 GCAGTGGAGATGCTTCAGATCACTTAAGGAGGCTTAATGAGGCGTAGC 5465
 425 aGlyLysAspAlaLysGlnIleAlaTyrIleLysAspLysLeuLeuH 442
 5466 AGCCTGTGTGGTACCAGAAATTTATGAAGGCCACAGACAGAGTCTAG 5515

442 isIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
 5516 ACCACCAATAGAGAAAGCATCAGCTGAGAAATAGCATTTGAAAGCTG 5565
 459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
 5566 AGAGCCCGC.....CTAGAAGCTGATGAAAGAAAGCA 5597
 475 sCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaL 492
 5598 G.....CTCTGTGTCTTCAACCAAC 5617
 492 euAsnGluMetTyrLysCysGlnAsnLeuLeuArgHisGlnValLysAsp 508
 5618 TGAAGGAAAGTGACCATCATCCAGATTACTTAAGGGTAGAGTGGAGAAC 5667
 509 Leu.....LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
 5668 CTTGAAAGAGAGCTAGAGATAGCCAGGACAAACCAAGAGCATGCAGCTCT 5717
 522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAsp 539
 5718 TGAGGCAGAGAAATTCCAAAGGAGAGGTAGAGACCTTAAAGCAAAATAG 5767
 539 roGlyLysAlaGlnAspPhe..... 545
 5768 AAGGATGACCCAAAGTCTGAGAGCTGGAAATAGATTGTTGTACTATA 5817
 546MetLysLysPheThrGlnValLeuGluAspAspGluLysI 559
 5818 AGGTCAGAAAGAAATCTGCACAAATCAATTACAAAAGAGCAAGAGCG 5867
 559 eArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnA 576
 5868 AATATCTGAATTAGAAATAATAATTC.....TCAT 5899
 576 laGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGln 592
 5900 TTGAAATATTTTGCAGAAAGAGCAAGAGAAAGTACAGATGAAGAA 5949
 593 ProThrAsnProPheLeuGluMet.....IleLysPheLeuLe 605
 5950 AAATCAAGCACTGCCATGGAGATGCTTCAACACAAATTTAAAGAGCTCAA 5999
 605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuI 622
 6000 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAGCCTGTAAAGGCC.... 6045
 622 leLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlu 638
 6046 ..AAGAGCAGAAATCTTAGT.....AGTCAAGTAGAG 6075
 639 GlyValProThrAspGlnAla...IleArgAlaGlyLeuGluLeuLeuLys 654
 6076 TGTCTTGAACCTTGAGAGGCTCAGTGTGTACAGGCTTGTATGAGGCCAA 6125
 654 sValLeuSerPheThrHisProIleSerPheHisSerAla..... 667
 6126 A.....AATAATTATATTGTTTTCATCTTCAATCTTCAAGTGAATGGCC 6163
 668GluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGlu 682
 6164 TCATTCAAGAGTAGAGTAGGCAAGCAAGCACTGGAGAAAGAGATGAA 6213
 683 LysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysI 699
 6214 GAAATCAGTAGACTGAAAAATCAAAT..... 6240
 599 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal.....L 714
 6241 ...CAAGACCAAGAGAGCTGTCTCTTAAACTGTCCAGAGGTGGAAGGAG 6286

1303LysGluGluProThrMetLysThrSer 1311
 3997 TACTCTCTTATCGCCAAAGATGTCGAGAGACCAATTTCGTGTCT 4046
 1311 rLysLysGlySerLysLysSerGlyProAlaProGluGlu... 1326
 4047 CTCCAAGAAGACTGAAGAATTC.....GAAGAGATGA 4081
 1327GluGluGluGluArgGlnSerGlyAsn 1336
 4082 TCATCAAGGTTATCGTCGGAGAACAAGAAATTGTTGCCAATCCAGAAC 4131
 1337 ThrGluGlnLysSerLysSerGlnHisArgValSerArg..... 1351
 4132 GAGGAGATCCCGCTTAAGGTTGACGGAAAGAAGATCCAATCTGAGGATTA 4181
 1352AlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerT 1367
 4182 TTCTGCTTACCAATCGAGAGACTCGGAGAGTCGCGTATTGTTATTGAGC 4231
 1367 hrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSer 1383
 4232 TTCCAGAAGGAGGCTTCGCTTGACGGATACACCATCAGACCACTT 4281
 1384 ProSerGlnProLysLysAsn 1390
 4282 CCATCTACAGCGCCAGAAT 4302

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ86851

seq_documentation_block:

ID AAQ86851 standard; DNA; 8789 BP.

XX

| | |
|----|---------------------------|
| AC | AAQ86851; |
| XX | |
| DT | 27-FEB-1996 (first entry) |

XX Human mitotin gene.
DE
XX
KW Cell cycle; M phase; mitotin; retinoblastoma; mitosis; cell growth;
KW inhibition; ss.

XX

OS Homo sapiens

•

PN WO9511309-A2.

XX

PD 27-APR-1995.

XX

PF 24-OCT-1994; 941

| | | |
|----|-------------------------------|--|
| XX | 22-OCT-1993; | 93US-0141239. |
| PR | | |
| XX | (TEXA) | UNIV TEXAS SYSTEM. |
| PA | | |
| XX | Lee W, | Zhu X; |
| PI | | |
| XX | | |
| XX | WPI; | 1995-170229/22. |
| DR | | |
| DR | P-PSDB; | AAR72826. |
| XX | | |
| PT | Purified mammalian protein | mitosin and agents that bind it and |
| PT | Inhibit its action - | used to promote cell growth or to inhibit cell |
| PT | division and/or proliferation | |

Claim 8; Fig 8B; 61pp; English.

AAQ86951 is the human mitotin gene. Mitotin is involved in the regulation of the mammalian mitotic cell cycle. Mitotin as with E2F-1 (see AAR7284) interacts with the retinoblastoma protein (the retinoblastoma tumor suppressor gene product). Mitotin is first synthesised at the G1/S boundary, it is then phosphorylated from S through M phase, and during mitosis, is closely associated with the centromeres/kinetochores at the mitotic spindle poles. Mitotin is necessary for a eukaryotic cell to enter the M phase of the mitotic cycle.

2532 CTCTGGAAGATCCCAACCGCTCATCTCGGCTGAAGGACAAATCTCTCTTG 2581
 784 roTPlYsSerTrp.....Val 789
 2582 AACTCGAGGAACGGAACCTCGTTGACCGTTGAGGCTCGCCCATCTGTT 2631
 790 AlaThrPheLeuValLysAspLeuLeuMetAsnAspArgLeu.....Pr 804
 2632 GCTGCCACCCAGCTCTACGAGATGAGAATGTTCACCCCACTCTTCGAGCA 2681
 804 oGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProG 821
 2682 AGGAGTCAAGAGCGTT.....CAATCGGTCGCGCCT 2713
 821 LuThrMetValLysLeuAlaLeuLysMetMetValArgTrpLeuLeu 837
 2714 ACACCCCAATCAAGATTCAGCGCTC.....GCT 2742
 838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLe 854
 2743 GGAATGAAGAGAACTTCGAA..... 2763
 854 uThrThrLeuHisSerAspGlyAspLeuThrGluGlnGlyLysLeu 871
 2764ATCGCTTAC.....AAGGTG 2779
 871 erLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaLeuVal 887
 2780 TCGTTCGAGAGAACCAAAA.....TCGATCGTC 2808
 888 LysLeuAlaGlnGluProCysTyr..... 895
 2809 TCCCTCACCCTCGCCAGTGTCTCTCTCGTTTCCAGGATTCGAA 2858
 896 HisGluLeuLeuThrLeuGluGlnThrGlnLeuCysAlaLeuAlaLea 912
 2859 GTTCGAATACATTGAGGCTGAAGAGAACCGCTGTTGTTCCCAATGGC 2908
 912 snAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLys 928
 2909 AACAAAGACTCAGGAATTCAGAAAGCTCTC..... 2940
 929 GlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaLeuCysAlaLe 945
 2940 2940
 945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysL 962
 2941AACTTC 2947
 962 euValLysAsnLeuAsnValArgArgGluTyrLeuLysGlnHisAla... 977
 2948 TCGGACTTGAAGTCTCCACCGCGGAACATCTCTCAACCAACACACTCTT 2997
 978AlaValSerGluLysLeuLe 984
 2998 GAAAACTGGCTTCTTCGAGCAAGACTTCGAGGTTTCGTTGAGAACA 3047
 984 uSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
 3048 GAACCGTCCAGCGCAATTCATGTCGCTCTCACCGTTGGACAACCTC... 3093
 1001 isAspProAspTyrValLysValGlnAspIleGluGlnLysAspVal 1017
 3094GAGAAAGCCGAACCTCTCCAAATC 3117
 1018 LysGluCysLeuTrpPheValLeuLeuIle...LeuMetAlaLysAsnGln 1033
 3118 AAGTACAACAAGATCTTCGAAAGAGAGTTCGAGCTCGAACAAGAGACAC 3167
 1033 uAsnAsnSerHisAlaPheLeuArgLysMetValGluAsnIleLysGlnT 1050
 3168 CGAGAGCGCGGCTGAGTACTTCAACAAGATGGTCAAGAACATCCAGAGG 3217

1050 hrLys.....AspAlaGlnGlyProAspAsp 1058
 3218 AGCAAGGATACAAATCGTCTATCTCTTGAACCTTGAGGCCCACTGAC 3267
 1059 AlaLysMetAsnGluLysLeuTyrThrValCysAsp..... 1070
 3268 TACACCATGAACACCGAGCTCACCAACGCTGTGTGACAAAGCAAGTCCGCAT 3317
 1071ValAlaMetAsnIleIleMetSerLysSerT 1081
 3318 GTGCCAATGGGAAGTTGAAATTCGCGCTCCCAATCTCTTGAGAAACCA 3367
 1081 hrThrThrSerLeuGluSer.....ProLysAspPro 1091
 3368 AGGATGGACTCTCCGCTCCCACTTCTTGTTCGCCCCAGAGATGCCA 3417
 1092 ValLeuProAlaArgPheThrGlnProAspLysAsnPheSerAsnT 1108
 3418 TCATCCCTCGTCAACTCCGCGATCAACCCACCGTGGAGTTC...AACT 3464
 1108 hrLysAsnTyrLeuPro.....ProGlu 1115
 3465 CTCTCTTACCTCTACCTGGGATCCCAAGAGAGCGAAAGTCCACGTC 3514
 1116 MetLysSerPheThrProLysProLysThrThrAsnValLeuGln 1132
 3515 ATGCTCAACTCCCAACATCCAAAGACAGAGAATAAGAGCGTAACATG 3564
 1132 yAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys 1149
 3565 GACCGTCAAT.....TCAATGGAATGCCAGAGTACGAGCTTT 3602
 1149 erSerArgMetGluThrValSerAsnAlaSer.....SerSerSer 1162
 3603 GATCAAGCGCGCGCTCTTAACCAAGATCAACGCTGTGCTGAGTACAAGC 3652
 1163 AsnProSerProGlyArgIleLysGlyArgLeuAspSerSerGluMe 1179
 3653 TCACCCGTGAACCCCAACAGTCTT.....GCCCGTTATTT 3689
 1179 tAspHisSerGluAsnGluAspTyr...ThrMetSerSerProLeuProG 1195
 3690 CGATCTTGTAAAGACTTACAACACTACTGACTGTCTCGAGC..... 3729
 1195 LysLysSerAspLysArgAspSerAspLeuValArgSerGluLeu 1211
 3730 ..CGCCCGAGACACACGAGAATATCGCGTCTGTCACCAACTCACAGTT 3777
 1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGlu... 1226
 3778 GAG...CCAATGTCGCCCAATACGTCAACATCACCATGCCCAATCCCAAT 3824
 1227 GluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGlnLysP 1243
 3825 GGAGCGCATTCAGCTCAAAAAGTT...CAAGTTCACGCTGTCTACCTTC 3871
 1243 roLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259
 3872 CATCCATGTCTCAACGTTCCGCTCAAG.....CACCAACTTACTGAGGCT 3915
 1260 AspGluGlnGlnTrpProGluGluLys.....ArgLeuLysGluAs 1273
 3916 TCCGGATCTGTCTCAAGGTCGAAAGAACCAACCAATCCGCACTTTCGATGA 3965
 1273 pIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysA 1290
 3966 TGTCTTTTACAAC..... 3978
 1290 rgGlyArgProProLysProLeuGlyGlyThrPro..... 1302
 3979ACTCCACTCACCACCTGC 3996


```

231 oTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThr... 246
   :: ||||| ::||| ::||| ::|||
1089 TTTCCGACACCTCCACCTCCCACTTAAGGAGATCCACGACTCTT 1138
   ::||| ::||| ::||| ::|||
247 .....SerIleSerAspLeuSerGluHisValPheAsp 257
   ::||| ::||| ::||| ::|||
1139 ACGTCAAGGCTGACAAGAGATCCAATCATGTGAGGACGCTCTT... 1185
   ||| ||| ::||| ::||| ::|||
258 LeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValle 274
   ||| ||| ::||| ::||| ::|||
1186 GCTATTGCTGGAACCAAGAACACAAATCAACACATCTTGTTCACATG.. 1233
   ::||| ::||| ::||| ::|||
274 uProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeu.. 290
   ::||| ::||| ::||| ::|||
1234 .....GAAACGAGGACATTCTCC 1252
   ::||| ::||| ::||| ::|||
291 .....GlnValValLysLeuLeuAlaLys... 298
   ||||| ::||| ::||| ::|||
1253 CACTCGGTCAATTCCTTAAGACCATCCAGACACACCATCCCATCCAA 1302
   ::||| ::||| ::||| ::|||
299 .....MetPheGlyAlaLysAspSerGluLeuAlaSerGI 310
   ::||| ::||| ::||| ::|||
1303 AGCATCGCTGAGGCACTCATCAAGTTCCGCGAGCTCGCGTGCACAGAA 1352
   ::||| ::||| ::||| ::|||
310 nAsnLysProLeuTrpGln..CysTyrLeuGlyArgPheAsnAspIleHi 326
   ||| ||| ::||| ::||| ::|||
1353 CAACCTGGTGTGCTGCAAGCTGCTGGTCCGCGCTGG..... 1391
   ::||| ::||| ::||| ::|||
326 sValProIleArgLeuGluCysValLysPheAlaSerHisCys...LeuM 342
   ::||| ::||| ::||| ::|||
1392 .....ATCAGTTGTCGCGG.....AATTGTTGACTAC 1419
   ::||| ::||| ::||| ::|||
342 eAsnHisProAsp..LeuAlaLysAsp.....LeuThrGI 353
   ::||| ::||| ::||| ::|||
1420 AAGAACATCCGTCCTCACTCGTCTGTGAGGACAAGCGTGAGCTCAAGGAA 1469
   ::||| ::||| ::||| ::|||
353 uTyrLeuLysValArgSerHisAspProGluGluAla...IleArgHisA 369
   ||| ||| ::||| ::||| ::|||
1470 GTTCCCTCCGCGTATTCACCAACATCAAGATGCTGAGACACCATAG 1519
   ::||| ::||| ::||| ::|||
369 sValIleValSerIleValThrAlaAlaLysLysAspIleLeuVal 385
   ::||| ::||| ::||| ::|||
1520 AGAAGATTTGGCCCTCAAGACCATCGGAACGCTGGACTC..... 1560
   ::||| ::||| ::||| ::|||
386 AsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArg.. 401
   ||| ||| ::||| ::||| ::|||
1561 ...GACATCTCCGCTCAACCAACTCAACGAGATCATTTGATAAGCGCCA 1607
   ::||| ::||| ::||| ::|||
402 ....TrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrL 417
   ||||| ::||| ::||| ::|||
1608 ACCATCCGAGTCCGCAAGGAGCCATGATGCCCTCAGACTCTT... 1653
   ::||| ::||| ::||| ::|||
417 yLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIle 433
   ||||| ::||| ::||| ::|||
1654 .....AAGACACCATGCCACG... 1671
   ::||| ::||| ::||| ::|||
434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAs 450
   ||||| ::||| ::||| ::|||
1672 ...AAGATCCAAAGAGTACTCTCCCAATCTACAAGACAGACAATACGA 1718
   ::||| ::||| ::||| ::|||
450 pAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisA 467
   ::||| ::||| ::||| ::|||
1719 GCCAGAGATC.....CGCATGCTTGTGTGAGAAATGATGCATA 1759
   ::||| ::||| ::||| ::|||
467 snLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
   ::||| ::||| ::||| ::|||
1760 CCGCCCGCAGAGAGTCCCTCCTC..... 1782
   ::||| ::||| ::||| ::|||
484 LeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAs 500
   ||||| ::||| ::||| ::|||
1783 .....GTCCAAGTGTCTCCCAATGGAGAAGGAGACCAA 1817
   ::||| ::||| ::||| ::|||
500 n.....LeuLeuArgHisGlnValLysAspLeuLeuAspLeuI 513
   ||| ||| ::||| ::||| ::|||

```

```

1818 CCAACAAGTTGCTGCTCTTACCCACCATGATCCGTCCTACTTTGCCATGT 1867
   ::||| ::||| ::||| ::|||
513 leLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal 529
   ::||| ::||| ::||| ::|||
1868 CCACCAACCCA...TGTTACCAACGTGTTGCTATCGTTTCTCTCAAAGTT 1914
   ||| ||| ::||| ::||| ::|||
530 MetValIleThrArgAsnLeuProAspPro..... 539
   ::||| ::||| ::||| ::|||
1915 CTTTCCTTCACTCGTTATCAACCACAAGAACAAATGATCGCTTCTCCTA 1964
   ::||| ::||| ::||| ::|||
540 .GlyLysAlaGlnAspPheMetLysLys..... 548
   ::||| ::||| ::||| ::|||
1965 CGCTCAACTTCCACTCTCTTCAAAACCTCTCTCTGGAGCTCAATTTG 2014
   ::||| ::||| ::||| ::|||
549 ..PheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGlu 564
   ||||| ::||| ::||| ::|||
2015 ACTTTCGCCCTATCTTCGAGAGAAGCTCTTCTTGCACAAAGGATCTTAC 2064
   ::||| ::||| ::||| ::|||
565 ValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValAr 581
   ::||| ::||| ::||| ::|||
2065 GCTTCTCTTGTGATCGCTCTCGGA.....GGAACCTGGA 2099
   ::||| ::||| ::||| ::|||
581 gGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheL 598
   ::||| ::||| ::||| ::|||
2100 CAAGTACTTTCGCCAAATCGGATCTCTCAGCAGACATGGACAATATG 2149
   ::||| ::||| ::||| ::|||
598 euGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAsp 614
   ::||| ::||| ::||| ::|||
2150 TTCAGATG.....GCTCTTGAGAACTC..... 2172
   ::||| ::||| ::||| ::|||
615 ThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGI 631
   ||||| ::||| ::||| ::|||
2173 ...GAAAGCCTC..... 2181
   ::||| ::||| ::||| ::|||
631 yThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgA 648
   ::||| ::||| ::||| ::|||
2182 .....GAGAGGAGTCGACCACTGTGTTCGCGGACGCGGTATTCAAA 2224
   ::||| ::||| ::||| ::|||
648 laGlyLeuGluLeuLysValLeuSer.....PheThr 659
   ::||| ::||| ::||| ::|||
2225 CTGGAATCAAGCTCTTAAGGAATTCGCCAGAAATGAACATCCGTGCT 2274
   ::||| ::||| ::||| ::|||
660 HisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCy 676
   ::||| ::||| ::||| ::|||
2275 CGCCAGCTACCTACACCGAGAGAGTGCCTCGCATGTGCTACCTCG 2324
   ::||| ::||| ::||| ::|||
676 sLeuLys...MetAsp.....AspGluLysValAlaG 686
   ||||| ::||| ::||| ::|||
2325 TTACAAGGACATGACTACGCTCTCTTCAATTCAGACAACTTGTGTG 2374
   ::||| ::||| ::||| ::|||
686 luAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAsp 702
   ||| ::||| ::||| ::|||
2375 AGAATCTTATCGAGAAGTTCACCAGCAACGGA...AAGGTTTCA... 2415
   ::||| ::||| ::||| ::|||
703 PheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLy 719
   ||| ::||| ::||| ::|||
2416 TTCAGCAAAATCGCGCTCTTCTC..... 2439
   ::||| ::||| ::||| ::|||
719 sLysGlyProProArgGluAlaLysTyrAlaIleHisCysIleHisAlaI 736
   ::||| ::||| ::||| ::|||
2440 .....AACCAAGAGTCGAGTTTCGAGACGCAC.....CATGCC.. 2472
   ::||| ::||| ::||| ::|||
736 lePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHis 752
   ||| ::||| ::||| ::|||
2473 .....GCTTACTTCTACGAAGCTATCCGC 2496
   ::||| ::||| ::||| ::|||
753 LysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuValTh 769
   ||||| ::||| ::||| ::|||
2497 AAATTC.....CCAACAACCTCTTGG.....CTTCCACTCACCAT 2531
   ::||| ::||| ::||| ::|||
769 rIleGlyHisIle.....AlaLeuLeuAlaProAspGlnPheAlaAlap 784
   ||| ||| ::||| ::||| ::|||

```



```

: |||::: |||::: ||| |||
9398 ACTTCCAGAAAGTTGTAAGAAAGGTTTGCTGACATCCCGACAGGAAAGA 9447
1300 lythrPro.....LysGluGluProThrMetLysThr.....Ser 1311
: |||::: |||::: ||| |||
9448 CTAGCCCATATATCTCTCGAAGAACCAACCACTGCGACAGCCCC 9497
1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluGlu 1328
: |||::: |||::: ||| |||
9498 CGCTGCTGCACAGAAAGTTAGCGCTATCCCTGAGTCTCGGCAAGA 9547
1328 uGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLys 1345
: |||::: |||::: ||| |||
9548 AATCTTGCAGAGTCTCTCAAAACCAACAGCTGGTGCGACAGATCACAA 9597
1345 lnHisArgValSerArgAlaGlnGlnArgAlaGluSerProGluSer 1361
: |||::: |||::: ||| |||
9598 AGGTCAAGTTGTCACGG.....AGCCCAAGTAGAT 9629
1362 SerAlaIle.....GluSerThrGlnSerThrProGlnLysGln 1374
: |||::: |||::: ||| |||
9630 TCAGGCACCATCTCCGAGAACCCACCAAGAAATCCGTCCTCCAGTCAATA 9679
1374 YArgGlyArgProSerLysThrProSerProSerGln 1386
: |||::: |||::: ||| |||
9680 TCTTCTGAGAGAAGTCCGACTGACAGCCCGCAGAG 9716

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA13177
seq_documentation_block:
ID AAA13177 standard; DNA; 4842 BP.
XX
AC AAA13177;
XX
DT 20-JUL-2000 (first entry)
XX
DE Vitellogenin encoding gene vit-2 nucleotide sequence.
XX
KW Vitellogenin; vit-2; primer set; detect; endocrine disturbance;
XX toxicity; environmental contamination; ds.
XX
OS Caenorhabditis elegans.
XX
PN JP2000069977-A.
XX
PD 07-MAR-2000.
XX
PF 03-SEP-1998; 98JP-0249723.
XX
PR 03-SEP-1998; 98JP-0249723.
XX
PA (KUMA-) KUNAMOTO KOTAI KENKYUSHO KK.
XX
DR WPI; 2000-264452/23.
XX
PT A primer for the amplification of nematode vitellogenin gene - for
PT evaluation of the endocrine disturbing activity of a chemical substance
XX
PS Claim 3; Page 10-12; 17pp; Japanese.
XX
CC This sequence represents the Caenorhabditis elegans vit-2 gene, which
CC encodes vitellogenin. The present invention relates to a primer set (see
CC AAA13179-A13188) containing primers which hybridise to different parts
CC of the common regions of the vit-1, vit-2, vit-3, vit-4 and vit-5 genes.
CC The vit-1 to 5 genes encode nematode vitellogenin. The invention also
CC includes a method for the detection of a vitellogenin gene in which a
CC gene encoding nematode vitellogenin is amplified using the primer set. A
CC method for evaluating the endocrine disturbing activity of a chemical
CC substance in which a gene encoding vitellogenin is amplified using the
CC primer set is also included in the invention. The method for evaluation
CC can be used for research on the toxicity of chemical substances and
CC environmental contamination.
XX

```

```

SQ Sequence 4842 BP; 1370 A; 1412 C; 1005 G; 1055 T; 0 other;

alignment_scores:
  Quality: 197.50      Length: 1576
  Ratio: 0.294        Gaps: 81
  Percent Similarity: 42.640  Percent Identity: 19.162

alignment_block:
US-09-512-581-2 x AAA13177 ..

Align seg 1/1 to: AAA13177 from: 1 to: 4842

81 AlaCysCysLeuAlaAspIlePheArg,IleTyrAlaPro..... 93
: |||::: |||::: ||| |||
414 AGCTGTGTCAACATGATCTCTTTCAACCCCAANTGCTCCAGAAACGAGA 463
94 .....GluAlaProTyrThrSerProAspLysLysLeuLysAspIle 106
: |||::: |||::: ||| |||
464 TTGAGAAGATCGAATCTCTTATGACAAGGAGGAACAATCTGAGGAAAC 513
107 PheMetPheIleThrArgGlnLeuLysGlyLeuGluAsp..... 119
: |||::: |||::: ||| |||
514 ACTTCTTTCTTCCACCAACGAG...AAGACCCCTTGAGGAGACTGCCAAGT 560
120 .....ThrLysSerP 123
: |||::: |||::: ||| |||
561 CGCTACACTGTCATCCGTGAGCAGAGAAGACCATCATCACCAGTCCA 610
123 roGlnPheAsnArg..... 127
: |||::: |||::: ||| |||
611 TCAACTTCGATAAGTGCACCGCAACGTTCCGAGATCGCTTACGGACTTCGT 660
127 ..... 127
661 TACTCTTCCGAGTCCCAAGATCGAGAAGGACACCGAGCTCATCCGCC 710
128 .....TyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerT 141
: |||::: |||::: ||| |||
711 ACAACTGTCTACACCTACGTTTGGAGAAC..... 741
141 yrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeu 157
: |||::: |||::: ||| |||
742 .....GAAGAGCTTAAGGAATCCGAG..... 762
158 TyrArgThrLeuPheSerValIleAsnAsnGly..... 168
: |||::: |||::: ||| |||
763 GTCGCTCCCTTTACACTGTCAACGTCACCGCAAGAAAGTAATGAAGAC 812
169 .....HisAsnGlnLysValH 174
: |||::: |||::: ||| |||
813 TGAACCCGCTCAAAAGTCGTTCTTGAGAGAACCACTCCATCAAGAGCC 862
174 isMetHisMetValAsp..... 179
: |||::: |||::: ||| |||
863 ACATCAGAAGGTCATGAGAGAGGAAGAACATCATCTACTCTTCCCGT 912
180 .....LeuMetSerSerIleCysGluGlyAspThrVal..... 191
: |||::: |||::: ||| |||
913 TGGGAGCAACTCGTTGAGGACTTCTCAAGAATGGAGCAAGGCTGAGTT 962
192 .....SerGlnGluLeuLeuAspT 198
: |||::: |||::: ||| |||
963 CGCCCCATTCGAGAAGTCCCATCTGCAAGAGAGATGCATCTGATCAAGA 1012
198 hrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAla 214
: |||::: |||::: ||| |||
1013 CTATCACTGAGCAATCCAGAAGTCGAGAACACATGCCAGAGACCTCC 1062
215 TyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluPr 231
: |||::: |||::: ||| |||
1063 CACTTCCTTGCTCGT.....CTTGTCCGCAT 1088

```


1470 1470
277 LeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVa 293
1471GAAGAAGAAAGATCAAGTGTT 1493
293 llyLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerG 310
1494 AGATTTAATC.....AAAGATGAAACCTA..... 1518
310 lnaAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
1519 ..AAAAAAGCTTA.....CTGAACCAACAAAA 1545
327 ValProIleArgLeuGluCysValLysPheAla..... 337
1546 GTTCAAGTGGCTAGATGTTTGAANAACGCTAAACCGATGAAGAAG 1595
338 .SerHisCys.....LeuMetAsnHisPro..... 345
1596 AAACGAGTGCCTAAAACTCATAAATGACCTTGAGATTAGAGAAATTC 1645
346AspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
1646 GTAAGGAATTAGAGCTTCAAAAAGAGCTTCAAGAGTAT..... 1683
359 SerHisAspProGluAlaIleArgHisAspValIleValSerIleVa 375
1684AAGATTGTATCAAAAACGCCAA 1706
375 lThrAlaAlaLysLysAspIleLeuLeu.....ValAsnAspHisLeuL 390
1707 AACAGAAGCTGAGAAAACAAATGCTTGAAGGCTTGCTAAAGAGCTA 1756
390 euAsnPheValArgGluArgThrLeuAsp.....LysArg 401
1757 TAGAGAGATTGAACAGCAGCAGCTAGATTGTTGAAAACGCTAAAAAC 1806
402 TrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLy 418
1807 GATGAAGAACGAACGAGTGCTTGAANAATATATCCCAAGACTTGCAAAA 1856
418 sTyrAlaLeuGln.....Sera 424
1857 AGAATATTAGCTGATATGAGCGTCAAGCGTTACAAGGATTGCGTATCAA 1906
424 laAlaGlyLysAspAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
1907 AAGCTAGAAATGAAAAGAGAACCAAGATGC.....GAGAAATTG 1947
441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluAr 457
1948 CTCACGCTGAAGCGGAAAAGTTAGAACACACAGGTTCTA..... 1989
457 gilePheAlaGlnTyrMetValProHisAsnLeuGluThr...ThrGluA 473
1990GATTGTTTGAANAACGCTAAACCCGATGAGAAC 2023
473 rGMetLysCysLeuTyrTyrLeuTyrAlaThrLeu.....AspLeuAsn 487
2024 GAAAAAGTGTTTGAAGATCTCCCTAAAGACTTACAAAGCGATATTCTA 2073
488 AlaValLysAlaLeuAsn..... 493
2074 GCCAAGAGAGCTGAAAGCTTTAAAGACTGCGTATCTCAAGCCAAAC 2123
494GluMetTrpLysCysGlnAsnLeu..... 502
2124 CGAAGCTGAGAAAAAAGATGGAGAATATTACTACCCCTCAAGCGGAAA 2173
503ArgHisGlnValLysAspLeuLeuAspLeu 512
.....GTGCTAGATTGTTTGAACACGCTAAA 2883


```

|||||
113 AGCATCCGTGATAAGGATGTCCTTACTGTAGTGGTGCCTGCTGCTGAT 64
87 IlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLysLe 103
|||||
63 ATTTTCAGGATTATGCTCCTGAGCTCCTTACACATCCCTGATAAACT 14
103 ulys 104
|||||
13 AAAG 10

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NAL1998.DAT:AAx14598

seq_documentation_block:

ID AAX14598 standard; DNA; 5334 BP.

```

XX AC AAX14598;
XX DT 31-MAR-1999 (first entry)
XX DE H. pylori GHPO 1755 gene.
XX KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX KW peptic ulcer disease; ss.
XX OS Helicobacter pylori.
XX FH Key
XX FT Location/Qualifiers
XX FT CDS
XX FT /*tag= a
XX PN W09843478-A1.
XX PD 08-OCT-1998.
XX PF 01-APR-1998; 98WO-US06371.
XX PR 29-JUL-1997; 97US-0902615.
XX PR 01-APR-1997; 97US-0833457.
XX PR 24-JUN-1997; 97US-0881227.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
XX DR P-PSDB; AAW98879.
XX CC New isolated Helicobacter polynucleotides - used to develop products
XX PT for the diagnosis, prevention and treatment of Helicobacter
XX PT infections and gastrointestinal diseases
XX PS Claim 1; Page 2000-2008; 2054pp; English.
XX CC This sequence represents a polynucleotide of the invention. It was
XX CC isolated from Helicobacter pylori and encodes a H. pylori GHPO protein.
XX CC The polypeptides can be used for preventing or treating Helicobacter
XX CC infections, and gastroduodenal diseases associated with these
XX CC infections, including acute, chronic, and atrophic gastritis, and peptic
XX CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
XX CC for the production of antibodies. The products can also be used for
XX CC detection and diagnosis.
XX SQ Sequence 5334 BP; 2206 A; 898 C; 1070 G; 1159 T; 1 other;

```

```

alignment_scores:
  Quality: 225.00      Length: 1545
  Ratio: 0.342        Gaps: 71
Percent Similarity: 42.524 Percent Identity: 18.447

```

alignment_block:

```

US-09-512-581-2 x AAX14598 ..
Align seg 1/1 to: AAX14598 from: 1 to: 5334
4 SerLysThrArgThrAsn...AspGlyLysIleThrTyrProGlyVa 19
|||||
955 ACAGAAACTCAACCAATTTTGTGGAGACAAGTCA..... 990
19 llysGluileSerAspLysIleSerLysGluGluMetValArgArgLeuL 36
|||||
991 .GAAGAATAACTGACGACTCAACGATCAAGAGATTATCAAGGAAGCA 1039
36 ys..... 36
1040 AAAGAAATATATATTGCTGGCATTGTAGTCGCTTCTTATCGTATT 1089
37 .....MetValLysThrPheMetAspMetAspGlnAs 48
1090 ATTTTATTTCTAGAAGCAATTTTCTACTACTTCATCGCTTGGAGATAA 1139
48 pSer.....GluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuH 62
|||||
1140 AAGCTCTCGTTTTCAGAAAGACAGAGAACTCTTTATGTCATGAATCC 1189
62 isLeuAlaSerAspPhe.....PheLeuLysHisProGlyLysAspVal 76
|||||
1190 AAATAAGCAAGAGATATAACCGATTGCTGAAAGACGGAATGAAAGGC 1239
77 ArgLeuLeuValAlaLacCysCysLeuAlaAspIlePheArgIleTyrAlaPr 93
|||||
1240 AATATGATCGAT..... 1251
93 oGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheI 110
|||||
1252 .....AAGATCTTTCTTC..... 1266
110 leThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
|||||
1267 .....AATGACGATCCCAATAGAAC 1287
127 ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnI 143
|||||
1288 TTATACAACTATTG.....AATATTGCA..... 1311
143 eCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgT 160
|||||
1312 .....GAAATTGAGCACAAAAACCGTTG.....AGAG 1339
160 hrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHis 176
|||||
1340 CCTTTTATGAATGTATTAGTAATGTGCGCACTATGAAGAATGTTTGAAG 1389
177 MetValAspLeuMetSerSerIlelleCysGluGlyAspThrValSerGl 193
|||||
1390 CTTATC..... 1395
193 nGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnL 210
|||||
1396 .....AAAGACAAAAAAC 1408
210 euAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLysArgThrAla 226
|||||
1409 TTCAAGATCAG.....ATGAAAGACTCTTA 1434
227 GlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMetLe 243
|||||
1435 GAGGCT.....TATAACGACTGCATCAAAA 1460
243 uGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleL 260
|||||
1461 TGCCCAAACT..... 1470
260 euGluLeuTyrAsnIleAspSerHisLeuLeuSerValLeuProGln 276

```


PN EP1033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 XX P-PSDB; AAG02811.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 1; SEQ ID 2815; 71pp + CD-ROM; English.
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX Sequence 295 BP; 106 A; 55 C; 61 G; 73 T; 0 other;

alignment_scores:
 Quality: 501.00 Length: 97
 Ratio: 5.165 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-512-581-2 x AAC02817 ..
 Align seg 1/1 to: AAC02817 from: 1 to: 295
 833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSe 849
 1 GTTCGATGGCTACTGTGGAATGAAATAATATCAGTAATCAGGAACCTC 50
 849 rThrLeuArgLeuLeuThrThrLeuHisSerAspGlyAspLeuThrG 866
 51 TACCTTAAGATTGCTAACACAAATATTGTCATAGTGAGACTTGAC 100
 866 luGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882
 101 AACAGGGGAAATAGTAAACCCAGATATGTCAGCTGTGAGACTTGCT 150
 883 GlySerAlaIleValLysLeuAlaGlnGluProCysThrHisGluIle 899
 151 GGGAGTGCTATTGTGAAGCTGGCACAAGAACCTTGTTACCATGAATCAT 200
 899 eThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys 916
 201 CACATTAGAACAATATCAGCTATGTGCATTAGCTATATCAACGATGAATGCT 250
 916 yrGlnValArgGlnValPheAlaGlnLysLeuHisLysGly 929
 251 ATCAAGTAAGACACAGTGTTTGGCCAGAACTTCACAAAGGC 291

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ80598
 seq_documentation_block:
 ID AAZ80598 standard; cDNA; 530 BP.
 XX AAZ80598;
 XX 07-APR-2000 (first entry)
 XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:682.
 XX Human; gene expression product; diagnosis; tumour; colon cancer;
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
 KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
 KW hyperplasia; ds.
 XX Homo sapiens.
 OS Homo sapiens.
 XX W09964576-A2.
 XX 16-DEC-1999.
 XX 09-JUN-1999; 99WO-IB01062.
 XX 10-JUN-1998; 98US-0088801.
 XX (FARB) BAYER CORP.
 PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
 PI Carroll E, Catino TU, Derti A, Ford DW, Lewis ME, Monahan JE;
 PI Schlegel R;
 XX WPI; 2000-087220/07.
 XX Novel nucleic acids, used to develop products for the diagnosis and
 PT treatment of disorders involving unwanted cell proliferation,
 PT particularly cancers, especially colon cancer -
 XX Claim 15; Page 402; 469pp; English.
 XX AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
 CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
 CC cDNA clones can be used to generate antisense oligonucleotides which
 CC can be used for antisense therapy. Methods and products from the present
 CC invention can be used for identifying and/or classifying cancerous cells
 CC present in a human tumour, particularly in solid tumours, e.g.
 CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
 CC can be used for developing agents for the diagnosis and treatment of
 CC disorders involving unwanted cell proliferation, such as neoplasia,
 CC dysplasia or hyperplasia.
 XX Sequence 530 BP; 191 A; 90 C; 83 G; 166 T; 0 other;
 alignment_scores:
 Quality: 349.00 Length: 68
 Ratio: 5.209 Gaps: 0
 Percent Similarity: 98.529 Percent Identity: 98.529
 alignment_block:
 US-09-512-581-2 x AAZ80598/rev ..
 Align seg 1/1 to reverse of: AAZ80598 from: 1 to: 530
 37 MetValLysThrPheMetAspMetAspGlnAspSerGluGluGly 53
 213 ATGCTTGTGAAACTTTTATGGATATGCACGAGACTCTGAAGAAGAAA 164
 53 sGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeu 70
 163 GGAGCTTTATTAAACCTAGCTTTACATCTGCCTTCAGATTTTTTCCTCA 114
 70 yHisProGlyLysAspValArgLeuValAlaCysCysLeuAlaAsp 86

xx
PS Claim 1; Page 126; 641pp; English.
xx
CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
xx
SQ Sequence 439 BP; 150 A; 101 C; 85 G; 103 T; 0 other;

alignment_scores:
Quality: 698.00 Length: 140
Ratio: 5.095 Gaps: 0
Percent Similarity: 97.857 Percent Identity: 97.857

alignment_block:

US-09-512-581-2 x AAV87629 ..

Align seg 1/1 to: AAV87629 from: 1 to: 439

1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProly 1089
114 ATGGCCTACATCATGCAAGAGTACTACATACAGTTTGGAAATCCCTAA 63
1089 sAspProValLeuProAlaArgPheThrGlnProAspLysAsnPheS 1106
64 AGACCCGGTACTACACAGCTCGTTCTTCACTCAACCTGACAAAGATTTCA 113
1106 eAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThrPro 1122
114 GTACACCAAAATTTATCTGCTCTCTGAAATCAATCAATTTTCACTCT 163
1123 GlyLysProLysThrAsnValLeuGlyAlaValAsnLysProLeuSe 1139
164 GGAACCTTAAACACCAATGTTCTAGAGCTGTATACAGCCACTTTC 213
1139 rSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVals 1156
214 ATCAGCAGCAAGCAATCTCAGACCAATCATCAGCAATGGAACCTGTA 263
1156 eAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGly 1172
264 GCAATGCAAGCAGCAGCTCAATCCCAAGCTCTCCTGGAAGATAAAGGG 313
1173 ArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMe 1189
314 AGGCTTGATGTTCTGAATGATGATCAGTCAAGTGAATGAGATTACCAAT 363
1189 tSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspL 1206
364 GTCCTTCACTTTGCGGGGAAAAAAGTGACAAAGAGAGACGACTCTGATC 413
1206 euValArgSerGluLeuGlu 1212
414 TTGTAAGTCTGAATCGGAG 433

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ15259

seq_documentation_block:

ID: AAZ15259 standard; cDNA; 738 BP.

xx
AC AAZ15259;

xx
DT 12-OCT-1999 (first entry)
xx
DE Human gene expression product cDNA sequence SEQ ID NO:2728.
xx
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
xx
OS Homo sapiens.
xx
PN WO9938972-A2.
xx
PD 05-AUG-1999.
xx
PF 28-JAN-1999; 99WO-US01619.
xx
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
xx
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
xx
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
xx
DR WPI; 1999-494092/41.
xx
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
xx
PS Claim 1; Page 1324; 2479pp; English.
xx
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ1532 to AAZ1779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 548 polynucleotide sequences given in AAZ1532 to AAZ1779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
xx
SQ Sequence 738 BP; 268 A; 145 C; 168 G; 150 T; 7 other;

alignment_scores:
Quality: 609.00 Length: 122
Ratio: 5.075 Gaps: 0
Percent Similarity: 98.361 Percent Identity: 95.902

alignment_block:
US-09-512-581-2 x AAZ15259 ..

Align seg 1/1 to: AAZ15259 from: 1 to: 738

alignment_block:

US-09-512-581-2 x AAF18338 ..

Align seg 1/1 to: AAF18338 from: 1 to: 1217

```

965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaValSerGI 981
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 AATATCAGTATACGACGAGGAATACATTAAGCAGACATCCTATGCTACTGA 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
981 uLysLeuSerLeuLeuProGluTyrValValProTyrThrIleHisL 998
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 GAATATATATCATCTGTTGCTCAATATGTAGTTCATACATGATTCACC 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
998 eUeuLaHisAspProAspTyrValLysValGlnAspIleGluInLeu 1014
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 TGCTAGCCCATGATCCAGATTTTACAAGATCAACAAGATGTTGATCAGCTT 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1015 LysAspValLysGluCysLeuTTPheValLeuGluIleLeuMetalaly 1031
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 CGTGATATCAAGAGATGCCCTATGCTGCTTGAAGTTTAAATGACAA 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1031 sAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleL 1048
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 GAATGAACAATAGCCATGCCCTTTATGAAGAAGATGGCAGACAAATCA 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1048 ySgInThrLysAspAlaGlnGlyProAspAlaLysMetAsnGluLys 1064
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 AGTTAACCCAGAGATGCCAGTCTCCAGATCAATCCCAAGACAAATGAAAA 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerth 1081
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 CTGTATACAGTATGTGATGGCTCTCTGTGTATATAATAGTAAAGATGC 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1081 rThrTyrSerLeuSerProLysAspProValLeuProAlaArgPhep 1098
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 TTTGTCGATGCAGATTCACCAAGGACCCAGTCTCCCAATGAATTTT 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1098 heThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 TTACACAACCTGAAAGGACTTCTGTAAACGATAAGAGTTATATTTTCAGAA 451
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1115 GluMetLysSerPhePheThrProGlyLysProLysThrAsnValle 1131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 GAGACAAGAGTACTCTGTAAACAGAAAGCCAAAGCCTGCTGGAGTACT 501
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1131 uGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrL 1148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 AGTGCAGTAAATAAGCTTTATCAGCAAGCGGAGGAACCCCTATGTTA 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1148 ySerSerArgMetGluThrValSer.....AsnAlaSerSerSer 1162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
552 GAAGCAGCTGCAGCTGAGACTGGAAGCAATATTAATGTAATTCAGAGCTG 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 AsnProSerSerProGlyArgIleLysGlyArg...LeuAspSerSerGI 1178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
602 AACCCCTTCAACCGGAATCGATCAAGGAACAGAGTTCAGAGCCAGCAGA 651
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1178 uMetAspHisSerGluAsnGluAsp.....TyrThrMetSerSerProL 1193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 AACTGGAGTTAGTGAATAATGAAGAGAACCCCTGTGAGGATTTATTCAGTCA 701
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1193 euProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
702 CACCTGTAAGAATATTGAC..... 721
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1210 GluLeuLysProArgGlyArgLysLysThrProValThrGluGlnGI 1226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 ..... 738
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1226 uGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGlnLysP 1243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
739 A.....ATTAATCTGATCAGGCTACC.....CAGGCA 767
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1243 roLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
768 ACATCAGCAGTACCGAGGAAGAAACACAGTACAGCAGCAGCTGGTGCA 817
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1260 AspGlu...GlnGlnTrpProGluGluLysArgLeuLysGluAspIleLe 1275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
818 GAGAAATATCCACAAACAAACAGATGAGAAA..... 847
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1275 uGluAsnGluAspGluGlnAsnSerPro...ProLysLysGlyLysArgG 1291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
848 .....GTAGATGATCGGACCTCCGCCCTTCCAAACCCAGGAGAG 890
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1291 lyArgProProLysProLeuGly...GlyGlyThrProLysGluGluPro 1306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891 GAGCTGAGCCCAAGTCTGAATCTCAGGCAATGCTACCAAAAATGATGAT 940
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 ThrMetLysThrSerLysLysGlySerLysLysLysSerGlyProAla 1323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 CTAATAAACCTATTAAACAGGGAAGAGAGA.....GCTGC 978
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1323 aProGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnL 1340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
979 AGTGGGTCAGGAGACCCCTGGGGGTTTGGAGCAGGTAAT..... 1018
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1340 ySerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnAla 1356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1019 .....GCCAAAGCACCCAAACTGCAAGATTTAGCCAAAAGGCAGCA 1060
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1357 GluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLy 1373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1061 CCAGCAGAAAGACMAATGTACTTACAAAGGTRAAATGCAATTTGCCAAGG 1110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1373 sGlyArgGlyArgProSerLysThrProSerProSerGlnProLys 1388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1111 GAGAAATGAAGCCCAACAGACAGCGCTNCCAGYTTNTGCAAAA 1156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /SIDSL/gcgdata/geneseq/NA1999.DAT:AAV87629

seq_documentation_block:

ID AAV87629 standard; cDNA; 439 BP.

XX AAV87629;

XX DT 12-FEB-1999 (first entry)

XX DE EST clone DY17.

XX KW Expressed sequence tag; secreted protein; hematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX OS Homo sapiens.

XX PN W09845437-A2.

XX PD 15-OCT-1998.

XX PF 10-APR-1998; 98WO-US06956.

XX PR 10-APR-1997; 97US-0837312.

XX PA (GEM) GENETICS INST INC.

XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 XX PI Racine LA, Spaulding V, Treacy M;

XX DR WPI; 1999-070078/06.

XX PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries

478 TyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnG1 494
152 TATTACTTATATAGTCTAGTTGGATCCAAATGCTGTAAAGCTCTCAACGA 201
494 uMetTrrpLysCysGlnAsnLeuLeuArgHISGlnValLysAspLeuLeuA 511
202 AATGTGGAGGTGCAGAACATGCTCGGATCCATGACGCGAATATPVG 251
511 splLeuileLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527
252 ATTTGCACAGCAGCCTACATCAGAGCTAACTGTTCTGCCATGTTTGA 301
528 LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAs 544
302 AAACCTGATCACCATAGCAAGAAATTTGCGCTGACCGCGGGAAGACAA 351
544 pPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgL 561
352 TTTTGTGAGAAATTTTACCAGGTTCTCGCGGATGATGAGAAATCTCGGT 401
561 ysgLInLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGlu 577
402 CTCAGTTGGAGTTTAAATAGCCCAACCTGTTCTTGCRAACAAGCAGAT 451
578 GlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProTh 594
452 ATTTGTGTGAGAAATAGCCCGAACTTGCRAAATCTTAAAGCAACCAAC 501
594 rAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProV 611
502 AAATCCTTTCTAGAGATGTCAAATTTCTGTTGGAAGAATCGCACCTG 551
611 alHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLys 627
552 TGCACATTTGATTCAGAAAGCATAAGTGCACTAGTGAATGATGAATAAG 601
628 SerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspG1 644
602 TCAATAGAGGGGACAGCAGATGATGAAGAGGGGTGAAGTCCAGATAC 651
644 nAlaIleArgAlaGlyLeuGluLeuLysValLeuSerPheThrHisP 661
652 AGCTATCCGTTCTAGGACTTGAACCTCTTAAGGTTCTGCTTTTACACATC 701
661 rotSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeu 677
702 CTACCTCGTTCCACTCTGCAGACATATAGTCTGTTTACAGTGCCTA 751
678 LysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAs 694
752 AGAATGGAGGATCACAAGCTAGCAGAAAGCTGCTATTCAAAATTTTAGAAA 801
694 nThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuL 711
802 TACAGGTCAAAATAGAAACACACACCTTCCCCAGATACGATCCACCTTA 851
711 euProValLeuHisLysLysSerLysGlyProArgGlnAlaLys 727
852 TTCCCATTTTACATCAAAAGACAGAGGGGTACTCCACACCAAGCAAAA 901
728 TyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPh 744
902 CAGGCTGTGCATGTATACAGCCCATATTCCAAATAAAGAAAGTCCAGCT 951
744 eAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuG 761
952 TGCACAGATTTTTCAGCCACTAGTAGGAGTCTGAATGCTGATGTCGCCG 1001
761 luhisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAla 777
1002 AACCACTTATACTCCATAGTTTTCATTGCGCCACATTTCTATGTAGCA 1051

778 ProAspGlnPheAlaAlaProTrrpLysSerTrrpValAlaThrPheIleVa 794
1052 CCAGATCAGTTTCTCCCAATGAAATCTGTAGTACCAAAATTTATTTGT 1101
794 lLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysL 811
1102 GAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAGAATGGAANAAC 1151
811 euTrrpValProAspGluGluValSerProGluThrMetValLysIleGln 827
1152 TGTGGTCTCCAGATGAAGAGTTTCCCTGAAGTACTAGCAAAAGGTACAG 1201
828 AlaIleLysMetMetValArgTrrpLeuLeuGlyMetLysAsnHisSe 844
1202 GCAATTAACCTTCTGTAAGGTGGCTGTTGGGTATGAAAAACAACCCAGTC 1251
844 rLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSerA 861
1252 TAATCTGCCAATTTCAACCTTCCGTTTATTCAGCGATGTTGGTTAGTG 1301
861 spGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
1302 AGGTCACCTGCAGAGCAAAAGAGGATCAGTAAATCTGATATGCTCTCGC 1351
878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCy 894
1352 TTGCGATTTAGCTGCTGCTAGTGCATATTAAGCTTGTCTCAGGAACCTTG 1401
894 sTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaI 911
1402 TTACCATGAATTTATTTACCCAGAACAGTTTTCAGCTCTGTGCACATTGTTA 1451
911 leAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHis 927
1452 TTAATGATGAGTGTACCAGTAAGGACAGATATTGCTCAGAAGCTGCAT 1501
928 LysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAl 944
1502 AAGGCACTTGTGAAGTTACTTCTCCATTTGGAGTATATGCGCATCTTTG 1551
944 aLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnC 961
1552 CTGTGCTGCCAAGATCTTGTGAAGGAGAGAGAGCAGCAGCAGCAGCAAT 1601
961 ysLeuValLysAsnIleAsnValArgGluTyrLeuLysGlnHisAla 977
1602 GTTTACTGAAAAATATCAGTATACGAGGGAATACATTAAAGCAGAACTCT 1651
978 AlaValSerGluLysLeuSerLeuProGluTyrValValProTy 994
1652 ATGGCTACTGAGAAATTTATATCAGTGTGCGCTGAATATGTAGTCCATA 1701
994 rThrIleHisLeuAlaHisAspProAspTyrValLysValGlnAspI 1011
1702 CATGATTCACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAGATG 1751
1011 leGluGlnLeuLysAspValLysGluCysLeuTrrpPheValLeuGluIle 1027
1752 TTGATCAGCTTCTGTATATCAAAAGAGTGCCTATGTTTCATGCTTGAAGTT 1801
1028 LeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVa 1044
1802 TTAATGACAAAGAAATGAAACCAATAGCCATGCTTTTATGAAGAAGATGCC 1851
1044 lGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysM 1061
1852 AGAACAATCAAGTTAACCCAGATGCCAGTCTCCAGATGAATCCCAAGA 1901
1061 eAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMet 1077
1902 CAAATCAAAACCTGATACAGTATGTGATGTGCTCTCTGCTGTATAAAT 1951
1078 SerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuPr 1094

1267 uLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerP 1284
|||||
3950 AAAGAGGCTCAAGAGATATATTAGAAATGAAGATGAACAGATAGTC 3999
|||||
1284 rOPrOLysLysGlyLysArgGlyArgProPrOLysProLeuGlyGlyGly 1300
|||||
4000 CGCCAAAAAGGGTAAAAGAGCGCCACCACCAAAACCTCTTGGTGGAGGT 4049
|||||
1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy 1317
|||||
4050 ACACCAAAAGAGAGCCCAACAAATGAATCTCTAAAAAGAGCAAAAA 4099
|||||
1317 sLysSerGlyProProAlaProGluGluGluGluGluGluGlns 1334
|||||
4100 AAAATCTGGACCTCCAGCACGAGGAGGAGGAAGAGAGAGAGAGACAAA 4149
|||||
1334 eRGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArg 1350
|||||
4150 GTGGAATACGGAACAGACAGTCCAAAAGCAACAGACCCGAGTCAAGG 4199
|||||
1351 ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh 1367
|||||
4200 AGAGCAGCAGCAGCAGCAGATCTCTCTGAATCTAGTCAATGAATCCAC 4249
|||||
1367 rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP 1384
|||||
4250 ACATCCACACCCACAGAAAGCAGGAGGAGACCATCAAAAACGCCATCAC 4299
|||||
1384 rOSerGlnProLysLysAsnVal 1391
|||||
4300 CATCAACCAACCAAAAAAATGTG 4322

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA47423

seq_documentation_block:

AC AAA47423 standard; DNA: 3957 BP.

XX 20-OCT-2000 (first entry)

XX Sequence encoding human neuron-associated protein.

XX Neuron associated protein; NEUP; neurological disorder; epilepsy;
KW Ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW Alzheimer's disease; Pick's disease; Huntington's disease;
KW dementia; Parkinson's disease; demyelinating disease; meningitis;
KW prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW peripheral nervous system; PNS; myopathy; schizophrenia;
KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW AIDS; Addison's disease; adult respiratory distress syndrome;
KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW Werner syndrome, trauma; human; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 140..2704
FT CDS
FT /*tag= a
FT /product= Neuron associated protein

XX WO200034477-A2.

XX 15-JUN-2000.

XX 10-DEC-1999; 99WO-0530408.

XX 11-DEC-1998; 98US-0210083.

PR 11-DEC-1998; 98US-9123456.
PR 09-FEB-1999; 99US-0119365.
PR 16-MAR-1999; 99US-0124687.
XX (INCY-).INCYTE PHARM INC.
XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
PI Lu DAM, Azimzai Y;
XX WPI; 2000-423423/36.
DR P-PSDB; AAB01382.
XX New human neuron-associated proteins and polynucleotides encoding them,
PT useful for diagnosis, treatment and prevention of cell proliferative
PT disorders including cancer, neuronal and neurological disorders
XX Claim 9; Page 130-131; 145pp; English.
XX Human neuron-associated proteins (NEUP) can be used for for
CC treating or preventing a disorder associated with decreased
CC expression or activity of NEUP. Antagonists of NEUP are useful for
CC treating or preventing disorder associated with increased expression
CC or activity of NEUP. NEUP or their fragments or derivatives are
CC useful for treating neurological disorder such as epilepsy, ischemic
CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC disease, Pick's disease, Huntington's disease, dementia and
CC Parkinson's disease. NEUPs are also useful for treating other
CC demyelinating diseases, bacterial and viral meningitis, prion
CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC metabolic diseases of the nervous system, neurofibromatosis, other
CC developmental disorders of the central nervous system, cerebral
CC palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC other neuromuscular disorders, peripheral nervous system disorders,
CC inherited, metabolic, endocrine, and toxic myopathies, mental
CC disorders including mood, anxiety and schizophrenic disorders, a cell
CC proliferative disorder such as actinic keratosis, arteriosclerosis,
CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
CC disease, adult respiratory distress syndrome, allergies, ankylosing
CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC complications of cancer, hemodialysis, and extracorporeal circulation,
CC viral, bacterial, fungal parasitic, protozoal, and helminthic
CC infections, and trauma. This sequence was given the Incyte ID no.
CC 2888437CB1.
XX Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;
SQ

alignment_scores:

Quality: 2932.50 Length: 977
Ratio: 3.693 Gaps: 16
Percent Similarity: 81.269 Percent Identity: 60.901

alignment_block:

US-09-512-581-2 x AAA47423 ..

Align seg 1/1 to: AAA47423 from: 1 to: 3957

428 AspAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTy 444
:::|||||
2 GAAGCTCAGAGAAAGTCAGCTGGATGAAGCAAACTTCTGCATATTTA 51
444 rTyrGlnAsnSerIleAspArgLysLeuValGluArgIlePheAlaG 461
|||||
52 TTATCAGAACAGCATTTGACGACAACTGTTGGTAGAGAAAATCTTGTCTC 101
461 lntYrMetValProHisAsnLeuGluThrGluArgMetLysCysLeu 477
|||||
102 AGTATCTTGTCCCCCACAACTCTGGAACAGAGAGAGAAATGAAATGCTTA 151

|||||
2100 GAACTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCAATTCCTGC 2149
667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspGluLysV 684
|||||
2150 TGAACACATTGAATCATTAAGTGGCTGCTGAAATGGATGATAAAG 2199
684 aAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
|||||
2200 TAGCAGAAGCTGCACATACAAATTTTCAAAACACAGGAAGCAAAATTTGAA 2249
701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLys 717
|||||
2250 GAGGATTTCCACATCAGATCAGCTTGCCTTCCTGTTTACATCACAA 2299
717 sSerLysLysGlyProProArgGlnAlaLysTyAlaIleHisCysIleH 734
|||||
2300 ATCTAAAAAAGGACCCCGCTCAACCAAAATATGCCATTGATATCC 2349
734 isAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750
|||||
2350 ATGCGATATTTCTAGTAAGAGACCCAGCTTTCACAGATATTTGAGCCT 2399
751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767
|||||
2400 CTGCATTAAGAGCTAGATCCAGCAACCTGGAACATCTCATACACCATT 2449
767 uValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlap 784
|||||
2450 GGTACTATTTGTCATATGCTCTCTTGCACCTGATCAATTTGCTGCTC 2499
784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
|||||
2500 CTTGGAATCTTGGGTAGCTACTTTCATGTGTAAGATCTTCTCATGAAT 2549
801 AspArgLeuProGlyLysThrThrLysLeuTrpValProAspGluGI 817
|||||
2550 GATCGCTTCCAGGAAAAGAACAACTAAACCTTGGGTCCAGATGAAGA 2599
817 uValSerProGluThrMetValLysIleGlnAlaIleLysMetMetVala 834
|||||
2600 AGTATCTCTGAGACAATGGTCAAAATTCAGGCTATTAATAATGATGGTTC 2649
834 rGrTrpLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850
|||||
2650 GATGGCTACTTGAATGAAAATAATACAGTAAATCAGGAACCTTCTACC 2699
851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGI 867
|||||
2700 TTAAGATTGCTAACACATATATGTCATATGATGGAGACTTGCACAGAACA 2749
867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlys 884
|||||
2750 GGGGAAAATTAGTAACACAGATATGTCACGTCTGACACTTGCCTGGGA 2799
884 erAlaIleValLysLeuAlaGlnLupProCystTyThrHisGluIleIleThr 900
|||||
2800 GTGCTATTGTGAAGCTGGCACAGAACCCTGTTTACCATTGAAATCATCACA 2849
901 LeuGluGlnTyGlnLeuCysAlaLeuAlaIleAsnAspGluCystTyrgI 917
|||||
2850 TTAGAACAATATCAGCTATGTCATTAAGTATGCAACGATGAATGCTATCA 2899
917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
|||||
2900 AGTAACACAAGCTTTGGCCAGAAAATCTCACAAAGGCCCTTTCCCGTTAC 2949
934 rgLeuProLeuGluTyTrpMetAlaIleCysAlaLeuCysAlaLysAspPro 950
|||||
2950 GGCCTCCACTTGAGTATATGGCAATCTGTCCTTTGTCGCAAAAGATCCT 2999
951 ValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
|||||

3000 GTAAGGAGAGAGAGCTCATGCTAGGCAATGTTTGGTGAAAAAATAATAA 3049
967 nValArgArgGluTyTrpLeuLysGlnHisAlaAlaValSerGluLysLeuL 984
|||||
3050 TGTAAAGCGGGAGTATCTCAAGCAGCATGCAGCTGTTAGTGAATAATAT 3099
984 euSerLeuLeuProGluTyTrpValValProTyThrIleHisLeuLeuAla 1000
|||||
3100 TGTCTCTTCTACAGAGTATGTTGTTCCATATACAAATTCACCTTTTGGCA 3149
1001 HisAspProAspTyTrpValLysValGlnAspIleGluGlnLeuLysAspVa 1017
|||||
3150 CATGACCCAGATATGTCAAAGTACAGGATATGAACTTAAGATGT 3199
1017 lLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluA 1034
|||||
3200 TAAAGAATGCTTGTGTTTGTCTGGAATATTAATGGCTAAAAATGAAA 3249
1034 sAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr 1050
|||||
3250 ATAACAGTCAAGCTTTTATCAGAAAGATGGTAAAAATATTAACAAACA 3299
1051 LysAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeuTyTrh 1067
|||||
3300 AAAGATGCCCAAGGACCATGATGCAAAATGAATGAAACCTGTACAC 3349
1067 rValCysAspValAlaMetAsnIleIleMetSerLysSerThrTyTrs 1084
|||||
3350 TGTGTGTGATGTTGCCATCAATATCATCATGTCAAAGAGTACTACATACA 3399
1084 erLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
|||||
3400 GTTGTGAATCTCTTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3449
1101 ProAspLysAsnPheSerAsnThrLysAsnTyTrpLeuProGluMetLys 1117
|||||
3450 CCTGACAGAAATTCAGTAAACACCAAAATATATCTCCCTCTGAAATGAA 3499
1117 sSerPhePheThrProGlyLysProLysThrAsnValLeuGlyAlav 1134
|||||
3500 ATCATTTTCTCTCTGGAACCTTAAACCAACCAATGTTCTAGGAGCTG 3549
1134 aAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer 1150
|||||
3550 TTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAATCATCA 3599
1151 ArgMetGluThrValSerAsnAlaSerSerSerSerSerAsnProSerPr 1167
|||||
3600 CGAATGGAACCTGTAAGCAATGCAACGACGAGCTCAAAATCCAAGCTCTCC 3649
1167 oGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA 1184
|||||
3650 TGAAGAATAAAGGGAGGCTTGATAGTCTGAAATGGATCACAGTGAA 3699
1184 snGluAspTyThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
|||||
3700 ATGAAGATTACACAATGCTCTTCCACTTTCGCGGGAAAAAAGTGACAAG 3749
1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyAr 1217
|||||
3750 AGACGACACTCTGATCTTGTAGGCTGTAATGGAGAACCTTAGAGGCAG 3799
1217 gLysLysThrProValThrGluGlnGluLysLeuGlyMetAspLys 1234
|||||
3800 GAAAAAACGCCGCTCACAGACAGGAGGAGAAATTAGGTATGGATGACT 3849
1234 euThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArg 1250
|||||
3850 TGACTAAGTTGTTACAGGAACAGAAACCTTAAAGGCGAGTCAGCGAAGTCGG 3899
1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGI 1267
|||||
3900 AAAAGAGGCCATACGGCTTCAGNAATCTGATGAACACAGCTGGCCTCAGGA 3949

51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67
|||||
300 GAAGAAAGAGAGCTTTATTTAAACCTAGCTTTACATCTTCTTCAGATTT 349
67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84
|||||
350 TTTTCTCAAGCATCTCGTAAAGATGTTCTGCTTACTGTAGCTGTCTGCC 399
84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
|||||
400 TTGCTGATATTTTCAGGATTTATGCTCTCTGAAGCTCTTACATCCCT 449
101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117
|||||
450 GATAAACTAAAGGATATATTTATGTTTATACAGACAGGTTGAAGGGCT 499
117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134
|||||
500 AGAGGATACAAAGAGCCCAATTCATAGTATTTTATTTACTTCAGA 549
134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuLeuAspSer 150
|||||
550 ACATGTGCTGGTCAAGTCATATAACATATGCTTTGAGTTAGAGATAGC 599
151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167
|||||
600 AATGAAATTTTACCCAGCTATACAGAACCTTATTTTCAGTTATAACAA 649
167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184
|||||
650 TGCGCAACATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTTA 699
184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
|||||
700 TTATTTGTGAAGGTGATACAGTGTCTCAGAGGCTTTTGGATACGGTTTA 749
201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLe 217
|||||
750 GTAATCTGGTACCTGCTCATAGAAATTTAAACAGCAACATATGATTT 799
217 uAlaLysAlaLeuLysArgThrAlaGlnAlaIleGluProTyrIleT 234
|||||
800 GGCAAGGCTTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTA 849
234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250
|||||
850 CCATTTTTTAAATCAGGTTCTGATGCTTGGGAAACATCTATCAGCGAT 899
251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSe 267
|||||
900 TTGTCAGAGCATGCTTTTGACTTAATTTTGGAGCTCTACAAATATGATAG 949
267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284
|||||
950 TCATTTGCTGCTCTGTTTACCCAGCTGTAATTTAAATTAAGAGCA 999
284 snAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
|||||
1000 ATGATAATAGGAGCGCTCAAGTTGTTTAACTTACTTGGCAAAATGTTT 1049
301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317
|||||
1050 GGGCAAGAGGATTCAGAAATGGCTTCTCAAAACAGGCCACTTTTGGCAGT 1099
317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334
|||||
1100 CTACTTGGCAGGTTTAAATGATATCCATGTACCAATCCGCGCTGGAATGTG 1149
334 alLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350
|||||
1150 TGAAATTTGTAGCATTTGTCTCATGAACCATCTCTGATTTAGCAAAAGAC 1199

351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
|||||
1200 TTAACAGAGTATCTTAAAGTGAGGTACATGACCTGAGGAGCTATTAG 1249
367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
|||||
1250 ACATGATGTTTATGTTGTCATAGTTACAGCTGCTAAAAAGGATATTCTTC 1299
384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
|||||
1300 TGTGCAATGATCATTACTTAAATTTTGTGAGAGAGAACATTAGACAAA 1349
401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417
|||||
1350 CGATGGAGAGTACGCAAGAGCCATGATGGGACTTCCCCAAATTTTATAA 1399
417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434
|||||
1400 GAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAG 1449
434 latrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450
|||||
1450 CATGGATCAAAAGACAAATTCGTACATATATTTATCAAAATAGTATTGAT 1499
451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467
|||||
1500 GATCGACTACTTGTGAAACGGATCTTTGCTCAATACATGTTCTCTCACAA 1549
467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484
|||||
1550 TTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTGTATGCCACAC 1599
484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
|||||
1600 TGGATTTAAATGCTGTAAGACATTTGAATGAAATGTGAAATGTCAAAAT 1649
501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLy 517
|||||
1650 CTGCTCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAAGCAACCCAA 1699
517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534
|||||
1700 AACAGATGCCAGTCTCAAGGCCATATTTTCAAAAGTGTGTTATTACAA 1749
534 rGsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550
|||||
1750 GAAATTTACCTGATCTCTGGTAAAGCTCAGGATTTTCATGAAGAAATTCACA 1799
551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567
|||||
1800 CAGGTGTTAGAGATGATGAGAAATTAAGAAAGCAGTTAGAGTACTTGT 1849
567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584
|||||
1850 TAGTCCAAACATGCTCTCTGCAAGCAGGCTGAAGGTTGTGCTGCTGAAATAA 1899
584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
|||||
1900 CTAAGAAATTTGGCAACCCCAACAGCCTACAAATCTCTTCTGGAAATG 1949
601 lLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSe 617
|||||
1950 ATCAAGTTTCTCTTGGAGAGATAGACCTGTGCACATAGATACCGAATC 1999
617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634
|||||
2000 TATCAGTGTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAG 2049
634 spAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
|||||
2050 ATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTT 2099
651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667


```

FT      exon      /number= 10
FT      /tag= j
FT      /number= 11
FT      1353..1504
FT      /tag= k
FT      /number= 12
FT      1505..1618
FT      /tag= l
FT      1619..1700
FT      /tag= m
FT      /number= 14
FT      1701..1749
FT      /tag= n
FT      /number= 15
FT      1750..1889
FT      /tag= o
FT      /number= 16
FT      1890..2005
FT      /tag= p
FT      /number= 17
FT      2006..2111
FT      /tag= q
FT      /number= 18
FT      2112..2272
FT      /tag= r
FT      /number= 19
FT      2273..2396
FT      /tag= s
FT      /number= 20
FT      2397..2555
FT      /tag= t
FT      /number= 21
FT      2556..2624
FT      /tag= u
FT      /number= 22
FT      2625..2761
FT      /tag= v
FT      /number= 23
FT      2672..2885
FT      /tag= w
FT      /number= 24
FT      2886..3090
FT      /tag= x
FT      /number= 25
FT      3091..3205
FT      /tag= y
FT      /number= 26
FT      3206..3338
FT      /tag= z
FT      /number= 27
FT      3339..3458
FT      /tag= aa
FT      /number= 28
FT      3459..3521
FT      /tag= ab
FT      /number= 29
FT      3522..3667
FT      /tag= ac
FT      /number= 30
FT      3668..3773
FT      /tag= ad
FT      /number= 31
FT      3774..4213
FT      /tag= ae
FT      /number= 32
FT      4214..4438
FT      /tag= af
FT      /number= 33
FT      4314..4408
FT      misc_signal
FT      /tag= ag
FT      /note= "Destabilizing signal"

```

```

FT      exon      4439..5337
FT      /tag= ah
FT      /number= 34
FT      5312..5317
FT      /tag= ai
FT      5333..5337
FT      /tag= aj
FT      /note= "Poly adenylation cleavage site"
PN      WO200050454-A1.
XX      31-AUG-2000.
XX
XX      24-FEB-2000; 2000WO-US04732.
XX
XX      24-FEB-1999; 99US-0121461.
XX      (TUFT ) TUFTS COLLEGE.
XX
XX      Soto AM, Sonnenschein C, Geck P, Szelei J;
PI      WPI: 2000-565451/52.
DR      P-PSDB; AAY94702.
XX
XX      New human androgen-induced tumor suppressor cDNA sequence termed
FT      'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
FT      diagnosis and treatment of prostate cancer -
XX
XX      Example 4; Fig 6; 152pp; English.
XX
XX      This invention relates to a human androgen-induced tumour suppressor
CC      cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
CC      located on chromosome 13 at position 13q12-13q. AS3 has a role in
CC      inhibiting cell proliferation and use as a marker for the efficient
CC      diagnosis and treatment of prostate cancer. The invention includes AS3
CC      cDNA and protein sequences, a vector comprising the cDNA sequence, a host
CC      cell transfected with the expression vector, and a method for producing
CC      an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC      cytostatic activity, and acts to suppress cell proliferation. The AS3
CC      gene is useful as a marker for the efficient diagnosis and treatment of
CC      prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC      antisense agents for sequence specific modulation of gene expression. The
CC      AS3 protein may be used in the treatment of disorders caused by aberrant
CC      modification or mutation of a gene encoding an AS3 protein, misregulation
CC      of the AS3 gene or aberrant post-translational modification of the AS3
CC      protein. This sequence represents the human AS3 cDNA sequence with an
CC      additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC      compared with the claimed AS3 cDNA sequence AAA28051.
XX
XX      Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;
SQ

```

alignment_scores:

Quality: 7193.00 Length: 1391
Ratio: 5.171 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-512-581-2 x AAA28052 ..

Align seg 1/1 to: AAA28052 from: 1 to: 5355

```

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrPropr 17
|||||
150 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATACATATCCGCC 199
17 oGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArg 34
|||||
200 TGGGGTCAAGGAAATATACATAAAATATCTAAAGAGGAGATGGTGAGAC 249
34 rgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
|||||
250 GATTAAGATGGTGTGAAAACCTTTTATGGATATGGACGAGACTCTGAA 299

```


| | | |
|---|---|------|
| 3966 | ACACCAAAAGAAGAGCCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAA | 4015 |
| 1317 | slyssSerGlyProProAlaProGluGluGluGluGluGluGluGluGlnS | 1334 |
| 4016 | AAAACTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAACACAA | 4065 |
| 1334 | erGlyAsnThrGluGlnLysSerLysLysGlnHisArgValSerArg | 1350 |
| 4066 | GTCGGAATAACGGAACAGAAAGTCCAAAGCAACAGCACCGAGTGTC | 4115 |
| 1351 | ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh | 1367 |
| 4116 | AGAGCACAGCAGAGACAGAAATCTCTCTGAATCTAGTGCAATTGAA | 4165 |
| 1367 | rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP | 1384 |
| 4166 | ACAGTCCACACACAGAAAGGAGGAGGAGACCATCAAACGCCATCAC | 4215 |
| 1384 | roSerGlnProLysLysAsnVal | 1391 |
| 4216 | CATCACACCAAAAAAATGTG | 4238 |
|
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28052 | | |
| seq_documentation_block: | | |
| ID | AAA28052 standard; cDNA: 5355 BP. | |
| XX | AAA28052; | |
| XX | | |
| DT | 01-DEC-2000 (first entry) | |
| XX | | |
| XX | Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4 | |
| XX | | |
| KW | Androgen-induced tumour suppressor; androgen shutoff gene 3 | |
| KW | chromosome 13q12-13q; cell proliferation inhibitor; prostat | |
| KW | diagnosis; treatment; cytostatic; human; ss. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PH | Location/Qualifiers | |
| FT | 85..130 | |
| FT | /tag= a | |
| FT | /number= 1 | |
| FT | 131..257 | |
| FT | /tag= b | |
| FT | /number= 2 | |
| FT | 150..4325 | |
| FT | /tag= c | |
| FT | /product= "AS3" | |
| FT | /note= "Androgen shutoff gene 3 protein" | |
| FT | 258..461 | |
| FT | /tag= d | |
| FT | /number= 3 | |
| FT | 462..548 | |
| FT | /tag= e | |
| FT | /number= 4 | |
| FT | 549..646 | |
| FT | /tag= f | |
| FT | /number= 5 | |
| FT | 647..773 | |
| FT | /tag= e | |
| FT | /number= 6 | |
| FT | 774..938 | |
| FT | /tag= f | |
| FT | /number= 7 | |
| FT | 939..995 | |
| FT | /tag= g | |
| FT | /number= 8 | |
| FT | 996..1111 | |
| FT | /tag= h | |
| FT | /number= 9 | |
| FT | 1112..1206 | |
| FT | /tag= i | |

401 ArgTrpArgValArgLysGluAlaLamMetMetGlyLeuAlaGlnIleTyrLy 417
 |||||
 1266 CGATGGAGAGTAGCGCAAGAAAGCCATGATGGGACTTGGCCAAATTTATAA 1315
 |||||
 417 slySTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434
 |||||
 1316 GAAATATGCTTTACATGTCAGCAGCTGGAAGAAGATGCTGCAAAACAGATAG 1365
 |||||
 434 latrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450
 |||||
 1366 CATGGATCAAGACAAATTTGCTACATATATATCAAAATAGTATTGAT 1415
 |||||
 451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467
 |||||
 1416 GATCGACTACTTGTGAACGGATCTTTGTGCTCAATACATGTTCTCCACAA 1465
 |||||
 467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484
 |||||
 1466 TTGAAGAACTACAGAACGGATGAATGCTTTATATTACTTGTATGCCACAC 1515
 |||||
 484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
 |||||
 1516 TGGATTTAAATGCTGTGAAGCATTCGAATGAAATGTGGAATGTCAAAAT 1565
 |||||
 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLeuIleLysGlnProLy 517
 |||||
 1566 CTGCTCCGACATCAAGTAAGGATTTGCTTGACTTGATTAAGCAACCCAA 1615
 |||||
 517 snhrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534
 |||||
 1616 ACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAA 1665
 |||||
 534 rGAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550
 |||||
 1666 GAAATTTACCTGATCCTGTGTAAGGCTCAGGATTTTCATGAAGAAATTCACA 1715
 |||||
 551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567
 |||||
 1716 CAGGTGTTAGAAGATGATGAGAAAATTAAGAAAGCAGTTAGAAAGTACTTGT 1765
 |||||
 567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584
 |||||
 1766 TAGTCCACATGCTCTCGACAGCAGCTGAGGTTGTGTGCGTGAATAA 1815
 |||||
 584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
 |||||
 1816 CTAAGAAGTTGGCAACCCCAACAGCCCTACAAATCCTTTCTCGGAAATG 1865
 |||||
 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSe 617
 |||||
 1866 ATCAAGTTTCTCTTGAGAGAGATAGCACCCTGTGCATATAGATACCGAATC 1915
 |||||
 617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634
 |||||
 1916 TATCAGTGTCTTATTAACAGTGAACAAATCAATAGATGAACAGCAG 1965
 |||||
 634 sPAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
 |||||
 1966 ATGATGAAGATGAGGTGTTTCAACTGTATCAAGCCATCAGACGAGTCTT 2015
 |||||
 651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667
 |||||
 2016 GAACGTCTTAAGGTACTCTCATTTACATATCCCATCTCATTTTCATTTCTGC 2065
 |||||
 667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684
 |||||
 2066 TGAACATTTTGAATCATTTACTGGCTTGTCTGAAAATGGATGATGAAGAAAG 2115
 |||||
 684 alaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
 |||||
 2116 TAGCAGAAGCTGCATACAAATTTTCAAAACACACAGGAAGCAAAATTTGAA 2165
 |||||

701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLy 717
 |||||
 2166 GAGGATTTTCCACACATCAGATCAGCTTGTCTTCTCTTGTATACATCA 2215
 |||||
 717 sSerLysLysGlyProArgGlnAlaLysTyrAlaIleHisCysIleH 734
 |||||
 2216 ATCTAAAAAGGAGCCCGCTCAAGCAAAATATGCCATTCATTTGTATCC 2265
 |||||
 734 isAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750
 |||||
 2266 ATGCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCT 2315
 |||||
 751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767
 |||||
 2316 CTGCTAAGAGCCTAGATCCAAAGCAACCTGGAACATCTCATACACCAT 2365
 |||||
 767 vAlThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlap 784
 |||||
 2366 GGTACTATTGGTCATATTTGCTCTCTTGCACCTGATCAATTTGCTGCTC 2415
 |||||
 784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
 |||||
 2416 CTTGGAATCTTTGGGTAGCTACTTTTCATTTGTGAAGATCTTCTCATGAAT 2465
 |||||
 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluG 817
 |||||
 2466 GATCGCTTCCAGGGAAAAAGACAACATAAACTTTGGGTTCCAGATGA 2515
 |||||
 817 vAlSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834
 |||||
 2516 AGTATCTCTGAGACAAATGTCAAAATTCAGGCTATTAAATGATGGTTC 2565
 |||||
 834 rgtTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850
 |||||
 2566 GATGGCTACTTGAATGAAAAATAATACACAGTAAATCAGGAACCTTCTACC 2615
 |||||
 851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluG 867
 |||||
 2616 TTAAGATTCTTAACACAATATTTGCTAGTAGTGGAGACTTGCAGAAC 2665
 |||||
 867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlyS 884
 |||||
 2666 GGGGAAAAATTAGTAACACAGATATGTCACGTCTGACACTTGTCTGGGA 2715
 |||||
 884 erAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
 |||||
 2716 GTGCTATTGTGAAGCTGGCACAAAGCCCTGTTTACCATTGAAATCATACA 2765
 |||||
 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG 917
 |||||
 2766 TTAGAACAATATCAGCTATGTGCATTTAGCTATCAACGATGAATGCTATCA 2815
 |||||
 917 vAlArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
 |||||
 2816 AGTAAGACAAGTGTGTCGCCAGAAACTTCACAAAGGCCCTTCCCGTTTAC 2865
 |||||
 934 rGLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950
 |||||
 2866 GGCCTTCCACTTGATATATGGAATCTGTGCCCTTTGTGCAAAAGATCCT 2915
 |||||
 951 ValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
 |||||
 2916 GTAAAGGAGAGAGAGCTCATGCTAGGCAATGTTTGGTGAATAATATAA 2965
 |||||
 967 vAlArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuL 984
 |||||
 2966 TGTAAGCGCGGAGTATCTCAACAGCAGCATGCAGCTGTTAGTGAATAATAT 3015
 |||||
 984 euSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
 |||||
 3016 TGTCTCTTCTACAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3065
 |||||
 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVa 1017
 |||||

PN WO200050454-A1.
 XX 31-AUG-2000.
 XX 24-FEB-2000; 2000WO-US04732.
 XX 24-FEB-1999; 99US-0121461.
 XX (TUFT) TUFIS COLLEGE.
 XX Soto AM, Sonnenschein C, Geck P, Szelei J;
 XX WPI; 2000-565451/52.
 DR P-PSDB; AAY94702.
 XX New human androgen-induced tumor suppressor cDNA sequence termed
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
 PT diagnosis and treatment of prostate cancer -
 XX Claim 1; Fig 1; 152pp; English.
 XX This invention relates to a human androgen-induced tumour suppressor cDNA
 CC sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
 CC on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
 CC proliferation and use as a marker for the efficient diagnosis and
 CC treatment of prostate cancer. The invention includes AS3 cDNA and protein
 CC sequences, a vector comprising the cDNA sequence, a host cell transfected
 CC with the expression vector, and a method for producing an AS3 polypeptide
 CC comprising culturing the transfected cells. AS3 has cytostatic activity,
 CC and acts to suppress cell proliferation. The AS3 gene is useful as a
 CC marker for the efficient diagnosis and treatment of prostate cancer. The
 CC AS3 nucleic acid molecule can be used as a source of antisense agents for
 CC sequence specific modulation of gene expression. The AS3 protein may be
 CC used in the treatment of disorders caused by aberrant modification or
 CC mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
 CC or aberrant post-translational modification of the AS3 protein. This
 CC sequence represents the human AS3 cDNA sequence.
 XX
 SQ Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;

alignment_scores:

Quality: 7193.00 Length: 1391
 Ratio: 5.171 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-512-581-2 x AAA28051 ..

Align seg 1/1 to: AAA28051 from: 1 to: 5271

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProPr 17
 66 ATGGCTCATTTCAAAGACTAGGACCAATCATGTGAAATAATCATATCGGCC 115
 17 oGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgA 34
 116 TGGGGTCAAGGAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGAC 165
 34 rgLeuLysMetValValLysThrPheMetAspMetAspGlnAnSpSerGlu 50
 166 GATTAAGATGGTTGTGAAACTTTTATGGATATGGACGAGGACTCTGAA 215
 51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67
 216 GAAGAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCCTCAGATTT 265
 67 ePheLeuLysHisProGlyLysAspValArgLeuValAlaCysCysL 84
 266 TTTTCTCAAGCATCTCGTAAAGATGTCGCTTACTGGTAGCCCTGCTGCC 315
 84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
 TTTTCTCAAGCATCTCGTAAAGATGTCGCTTACTGGTAGCCCTGCTGCC

316 TTGCTGATATTTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCT 365
 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117
 366 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGCT 415
 117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134
 416 AGAGGATACAAAGAGCCACAATTCATAGGTATTTTATTACTTGAGA 465
 134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150
 466 ACATTGCTTGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGATAGC 515
 151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167
 516 AATGAATTTTACCCAGCTATACAGAACCTTATTTTCAGTTATAAACAA 565
 167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184
 566 TGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTA 615
 184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
 616 TTATTTCTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGTTTAA 665
 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLe 217
 666 GTAATCTGTGATCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTT 715
 217 uAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleT 234
 716 GGCNAAGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTA 765
 234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250
 766 CCACCTTTTTTAAATCAGGTTCTGATGCTTGGGAAACATCTATCAGCGAT 815
 251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSe 267
 816 TTGTGACAGCATGCTTTTGACTTAAATTTGGAGCTCTACAAATATGTAG 865
 267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284
 866 TCATTGCTGCTCTCTCTTTTACCCAGCTGAATTTAAATTAAGAGCA 915
 284 snAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
 916 ATGATAATGAGGAGCGCTACAAGTTGTTAAACTACTTGGCAAAATGTTT 965
 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317
 966 GGGCAAGGATTCAGAAATTTGGCTTCTCAAAACAAGCCACTTTGGCAGTG 1015
 317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334
 1016 CTACTTGGGAGGTTTAAATGATATCCATGTACCAATCCGCTCGAATGTG 1065
 334 allYsPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350
 1066 TGAATTTTGCTAGCCATTGCTCATGAACCATCTGTATTTAGCAAAAGAC 1115
 351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
 1116 TTAACAGAGTATCTTAAAGTGAGTGCATGACCTGAGGAGGATATTAG 1165
 367 qHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
 1166 ACATGATGTTATTGTGTCAATAGTTACAGCTGCTTAAAGGATATTCTTC 1215
 384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
 1216 TGGTCAATGATCACTTACTTAAATTTTGTGTGAGAGAGAGAACATTAGACAA 1265


```

/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT72157 + 129.00 142.91 5.45 2126
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAD00750 + 129.00 135.99 13.24 4193
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF54895 + 129.00 135.53 14.04 4386
/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT05406 + 129.00 135.33 14.40 4472
/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT16231 + 129.00 133.69 17.78 5254
/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:AAO79377 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV82916 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA95039 + 129.00 133.16 19.05 5538
/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:AAZ38730 + 129.00 130.08 28.27 7492
/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV25812 + 129.00 130.08 28.27 7493

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA28051
seq_documentation_block:
ID AAA28051 standard; cDNA; 5271 BP.
XX
AC AAA28051;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX
KW Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW diagnosis; treatment; cytostatic; human; ss.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT exon 1..46
FT FT /*tag= a
FT FT /number= 1
FT FT 47..173
FT FT /*tag= b
FT FT /number= 2
FT FT 66..4241
FT FT /*tag= c
FT FT /product= "AS3"
FT FT /note= "Androgen shutoff gene 3 protein, the CDS is
FT FT specifically claimed as SEQ ID #3"
FT FT
FT exon 174..377
FT FT /*tag= d
FT FT /number= 3
FT FT 378..464
FT FT /*tag= e
FT FT /number= 4
FT FT 465..562
FT FT /*tag= f
FT FT /number= 5
FT FT 563..689
FT FT /*tag= e
FT FT /number= 6
FT FT 690..854
FT FT /*tag= f
FT FT /number= 7
FT FT 855..911
FT FT /*tag= g
FT FT /number= 8
FT FT 912..1027
FT FT /*tag= h
FT FT /number= 9
FT FT 1028..1122
FT FT /*tag= i
FT FT /number= 10
FT FT 1123..1268
FT FT /*tag= j
FT FT /number= 11
FT FT 1269..1420
FT FT /*tag= k
FT FT /number= 12
FT FT 1421..1534
FT FT /*tag= l
FT FT
FT exon 1535..1616
FT FT /*tag= m
FT FT /number= 14
FT FT 1617..1665
FT FT /*tag= n
FT FT /number= 15
FT FT 1666..1805
FT FT /*tag= o
FT FT /number= 16
FT FT 1806..1921
FT FT /*tag= p
FT FT /number= 17
FT FT 1922..2027
FT FT /*tag= q
FT FT /number= 18
FT FT 2028..2188
FT FT /*tag= r
FT FT /number= 19
FT FT 2189..2312
FT FT /*tag= s
FT FT /number= 20
FT FT 2313..2471
FT FT /*tag= t
FT FT /number= 21
FT FT 2472..2540
FT FT /*tag= u
FT FT /number= 22
FT FT 2541..2677
FT FT /*tag= v
FT FT /number= 23
FT FT 2678..2801
FT FT /*tag= w
FT FT /number= 24
FT FT 2802..3006
FT FT /*tag= x
FT FT /number= 25
FT FT 3007..3121
FT FT /*tag= y
FT FT /number= 26
FT FT 3122..3254
FT FT /*tag= z
FT FT /number= 27
FT FT 3255..3374
FT FT /*tag= aa
FT FT /number= 28
FT FT 3375..3437
FT FT /*tag= ab
FT FT /number= 29
FT FT 3438..3583
FT FT /*tag= ac
FT FT /number= 30
FT FT 3584..3689
FT FT /*tag= ad
FT FT /number= 31
FT FT 3690..4129
FT FT /*tag= ae
FT FT /number= 32
FT FT 4130..4354
FT FT /*tag= af
FT FT /number= 33
FT FT 4320..4324
FT FT /*tag= ag
FT FT /note= "Destabilizing signal"
FT FT 4355..5253
FT FT /*tag= ah
FT FT /number= 34
FT FT 5228..5233
FT FT /*tag= ai
FT FT 5249..5253
FT FT /*tag= aj
FT FT /note= "Poly adenylation cleavage site"
FT FT
XX

```


| | | | | | | | | | |
|---|--------|--------|---------|---------|---|--------|--------|--------|------|
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AA760300 + | 132.00 | 140.12 | 7.79 | 4040 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC51256 + | 130.00 | 141.96 | 6.16 | 2640 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76440 + | 132.00 | 138.97 | 9.03 | 4522 | /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA10247 + | 130.00 | 139.81 | 8.11 | 3259 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA15096 + | 132.00 | 138.02 | 10.21 | 4967 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA15623 + | 130.00 | 139.36 | 6.59 | 3405 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAV02301 + | 132.00 | 137.80 | 10.50 | 5074 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV61587 + | 130.00 | 138.71 | 9.35 | 3632 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.AA050831 + | 132.00 | 133.75 | 17.66 | 7555 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAV03326 + | 130.00 | 136.31 | 12.71 | 4596 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.AA091779 + | 132.00 | 131.39 | 23.89 | 9520 | /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.AAO27189 + | 130.00 | 136.14 | 12.99 | 4673 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT45424 + | 132.00 | 131.28 | 24.22 | 9620 | /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.AAC76615 + | 130.00 | 131.94 | 22.26 | 7056 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC36211 + | 132.00 | 128.06 | 36.61 | 13198 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76473 + | 130.00 | 130.09 | 28.21 | 8458 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC36212 + | 132.00 | 125.71 | 49.48 | 16622 | /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.AAO20602 + | 130.00 | 129.53 | 30.32 | 8937 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAT06769 + | 132.00 | 120.18 | 100.55 | 28598 | /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAO50185 + | 130.00 | 129.45 | 30.64 | 9009 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT89956 + | 132.00 | 120.06 | 102.20 | 28958 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV25810 + | 130.00 | 129.45 | 30.64 | 9009 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA75299 + | 132.00 | 120.06 | 102.20 | 28958 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV12112 + | 130.00 | 129.45 | 30.64 | 9009 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA46806 + | 132.00 | 119.74 | 106.47 | 29879 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX91162 + | 130.00 | 129.45 | 30.64 | 9009 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AA260606 + | 132.00 | 115.87 | 174.85 | 43676 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAC90508 + | 130.00 | 128.45 | 30.64 | 9009 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV05287 + | 132.00 | 114.62 | 205.25 | 49377 | /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.AAO55134 + | 130.00 | 128.18 | 36.06 | 1020 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV68070 + | 131.50 | 145.50 | 3.91 | 2241 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT46941 + | 130.00 | 127.69 | 38.39 | 1070 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA28521 + | 131.50 | 145.46 | 3.93 | 2250 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA82261 + | 130.00 | 126.66 | 43.81 | 1184 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA20288 + | 131.50 | 142.56 | 5.70 | 2291 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV18277 + | 130.00 | 124.11 | 60.78 | 1521 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14958 + | 131.50 | 142.45 | 5.78 | 3024 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV18279 + | 130.00 | 124.11 | 60.78 | 1521 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT97611 + | 131.50 | 140.58 | 7.35 | 3633 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAC22911 + | 130.00 | 124.11 | 60.78 | 1521 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.AAO58708 + | 131.50 | 140.32 | 7.60 | 3728 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAC22913 + | 130.00 | 124.11 | 60.78 | 1521 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.AAN90489 + | 131.50 | 137.21 | 11.32 | 5056 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX35269 + | 130.00 | 124.11 | 60.78 | 1521 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA87968 + | 131.50 | 136.90 | 11.79 | 5214 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX35271 + | 130.00 | 124.11 | 60.78 | 1521 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14824 + | 131.50 | 136.90 | 11.79 | 5214 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14651 + | 130.00 | 120.77 | 93.30 | 2112 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT68770 + | 131.50 | 131.28 | 24.21 | 9046 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14651 + | 130.00 | 107.52 | 510.16 | 7753 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAT68761 + | 131.50 | 131.28 | 24.21 | 9046 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX88907 + | 129.50 | 154.85 | 1.18 | 700 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81502 + | 131.50 | 111.10 | 322.56 | 65632 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX39869 + | 129.50 | 154.07 | 1.30 | 756 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAT21544 + | 131.50 | 94.04 | 2.9e+03 | 349980 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC53422 + | 129.50 | 146.28 | 3.54 | 1624 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81430 + | 131.50 | 79.65 | 1.7e+04 | 1437668 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC49797 + | 129.50 | 146.09 | 3.62 | 1634 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA6314 + | 131.00 | 151.78 | 1.85 | 1138 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC42645 + | 129.50 | 146.07 | 3.63 | 1657 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC69638 + | 131.00 | 149.84 | 2.24 | 1377 | /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAO87587 + | 129.50 | 145.90 | 3.72 | 1686 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAT31963 + | 131.00 | 148.37 | 2.71 | 1590 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV42650 + | 129.50 | 142.61 | 5.66 | 2328 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX83421 + | 131.00 | 145.53 | 3.89 | 2101 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70260 + | 129.50 | 139.88 | 8.05 | 3045 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA50826 + | 131.00 | 145.53 | 3.89 | 2101 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT984930 + | 129.50 | 139.65 | 8.28 | 3114 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA83420 + | 131.00 | 145.26 | 4.03 | 2158 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA58860 + | 129.50 | 139.65 | 8.28 | 3114 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA50285 + | 131.00 | 144.26 | 4.03 | 2158 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV22251 + | 129.50 | 139.06 | 8.94 | 3300 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV42798 + | 131.00 | 145.45 | 4.27 | 2336 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV17259 + | 129.50 | 138.48 | 9.62 | 3300 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC75869 + | 131.00 | 142.50 | 5.75 | 2830 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV11348 + | 129.50 | 138.48 | 9.62 | 3300 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT72156 + | 131.00 | 142.44 | 5.79 | 2846 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC936459 + | 129.50 | 138.20 | 9.98 | 3591 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX78193 + | 131.00 | 141.77 | 6.31 | 3040 | /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.AAN91475 + | 129.50 | 138.07 | 10.15 | 3637 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX24135 + | 131.00 | 141.77 | 6.31 | 3040 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV11344 + | 129.50 | 136.60 | 12.25 | 4200 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA57868 + | 131.00 | 141.66 | 6.40 | 3073 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA12452 + | 129.50 | 136.60 | 12.25 | 4200 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAT958616 + | 131.00 | 139.79 | 8.13 | 3691 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC48567 + | 129.50 | 136.12 | 13.02 | 4402 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT976110 + | 131.00 | 139.35 | 8.60 | 3854 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT61865 + | 129.50 | 133.19 | 18.96 | 5868 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC51553 + | 131.00 | 139.04 | 8.96 | 3975 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA60104 + | 129.50 | 133.19 | 18.96 | 5868 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAO82834 + | 131.00 | 138.49 | 9.61 | 4193 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA62948 + | 129.50 | 130.93 | 25.34 | 7326 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAT37097 + | 131.00 | 137.96 | 10.28 | 4416 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA15554 + | 129.50 | 130.93 | 25.34 | 7326 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70078 + | 131.00 | 136.43 | 12.52 | 5136 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV70401 + | 129.50 | 123.32 | 67.23 | 1546 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA45239 + | 131.00 | 134.08 | 16.91 | 6463 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAO72121 + | 129.50 | 122.63 | 73.48 | 1655 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT85319 + | 131.00 | 133.58 | 18.04 | 6791 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAF61281 + | 129.50 | 111.72 | 297.82 | 4830 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT861548 + | 131.00 | 130.70 | 26.10 | 9009 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAC40067 + | 129.00 | 146.39 | 3.49 | 1511 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT89256 + | 131.00 | 129.47 | 40.52 | 12616 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAT61725 + | 129.00 | 143.58 | 5.00 | 1990 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA88744 + | 131.00 | 125.36 | 51.80 | 15225 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV62142 + | 129.00 | 143.25 | 5.22 | 2056 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAT18276 + | 131.00 | 125.35 | 51.82 | 15229 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAT08979 + | 129.00 | 143.11 | 5.31 | 2085 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAC22910 + | 131.00 | 125.35 | 51.82 | 15229 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAT07178 + | 129.00 | 143.11 | 5.31 | 2085 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAC35268 + | 131.00 | 125.35 | 51.82 | 15229 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV33264 + | 129.00 | 143.11 | 5.31 | 2085 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV30458 + | 131.00 | 89.10 | 5.4e+03 | 534720 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAZ25063 + | 129.00 | 143.11 | 5.31 | 2085 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV30459 + | 131.00 | 89.07 | 5.4e+03 | 536165 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX02012 + | 129.00 | 143.11 | 5.31 | 2085 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA33406 + | 130.50 | 151.23 | 1.87 | 1129 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAV84069 + | 129.00 | 143.11 | 5.31 | 2085 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA34456 + | 130.50 | 151.23 | 1.87 | 1129 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA31092 + | 130.50 | 144.24 | 2.07 | 1218 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAT92416 + | 130.50 | 140.46 | 4.60 | 2244 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV68069 + | 130.50 | 143.68 | 4.94 | 2370 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAV44303 + | 130.50 | 143.30 | 5.19 | 2460 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC45296 + | 130.50 | 139.42 | 8.53 | 3399 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA82833 + | 130.50 | 137.52 | 10.88 | 4337 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAT27052 + | 130.50 | 136.93 | 11.74 | 4599 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAT54902 + | 130.50 | 136.85 | 11.86 | 4632 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81487 + | 130.50 | 112.43 | 271.84 | 50925 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAT21610 + | 130.50 | 92.79 | 3.4e+03 | 349980 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA02346 + | 130.00 | 148.84 | 2.55 | 1343 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA47428 + | 130.00 | 144.92 | 4.21 | 1974 | | | | | |

| | | | | | | | | | |
|--|--------|--------|---------|--------|---|--------|--------|---------|------|
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AA70071 + | 136.00 | 141.60 | 6.45 | 5709 | /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AA86743 + | 134.00 | 144.91 | 4.22 | 3228 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV60567 + | 136.00 | 141.60 | 6.45 | 5709 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV70099 + | 134.00 | 143.86 | 4.83 | 3579 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1996.DAT:AA717480 + | 136.00 | 141.25 | 6.74 | 5910 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV61586 + | 134.00 | 143.71 | 4.92 | 3632 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ49991 + | 136.00 | 135.30 | 14.46 | 10594 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV70204 + | 134.00 | 141.95 | 6.17 | 4317 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ49992 + | 136.00 | 135.30 | 14.47 | 10598 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV61585 + | 134.00 | 141.72 | 6.35 | 4414 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AA868936 + | 136.00 | 133.93 | 17.24 | 12119 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV61678 + | 134.00 | 141.68 | 6.39 | 4434 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ49988 + | 136.00 | 133.62 | 17.34 | 12194 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV61677 + | 134.00 | 141.56 | 6.48 | 4485 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ49999 + | 135.50 | 160.46 | 0.5737 | 843 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV60421 + | 134.00 | 141.14 | 6.84 | 4674 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ49998 + | 135.50 | 152.29 | 1.64 | 1881 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV23611 + | 134.00 | 140.15 | 7.74 | 5137 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ49997 + | 135.50 | 151.27 | 1.86 | 2078 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV23610 + | 134.00 | 136.75 | 12.01 | 7192 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV02016 + | 135.50 | 149.29 | 2.41 | 2525 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV23634 + | 134.00 | 132.08 | 21.88 | 1137 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ00313 + | 135.50 | 149.28 | 2.41 | 2527 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV23633 + | 134.00 | 132.67 | 33.86 | 1589 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV71078 + | 135.50 | 147.56 | 3.00 | 2991 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ22970 + | 134.00 | 128.67 | 33.86 | 1589 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ27622 + | 135.50 | 145.97 | 3.68 | 3496 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ22971 + | 134.00 | 108.40 | 456.00 | 1162 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ27621 + | 135.50 | 145.22 | 4.06 | 3766 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV54649 + | 133.50 | 157.65 | 0.8228 | 869 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV54648 + | 135.50 | 143.59 | 5.00 | 4417 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV20215 + | 133.50 | 157.65 | 0.8228 | 869 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV54680 + | 135.50 | 143.59 | 5.00 | 4417 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV20218 + | 133.50 | 157.65 | 0.8228 | 869 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV54686 + | 135.50 | 143.59 | 5.00 | 4417 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV54093 + | 133.50 | 157.65 | 0.8228 | 869 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ34632 + | 135.50 | 143.59 | 5.00 | 4417 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV54096 + | 133.50 | 157.65 | 0.8228 | 869 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AAZ34631 + | 135.50 | 143.46 | 5.08 | 4473 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV59583 + | 133.50 | 149.04 | 2.48 | 2024 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ27618 + | 135.50 | 142.46 | 5.78 | 4936 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV59582 + | 133.50 | 148.43 | 2.68 | 2148 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ50918 + | 135.50 | 141.52 | 6.52 | 5415 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV16098 + | 133.50 | 145.53 | 3.90 | 2857 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV52328 + | 135.50 | 139.36 | 8.60 | 6693 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV52327 + | 133.50 | 145.43 | 3.95 | 2886 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV10093 + | 135.50 | 138.44 | 9.68 | 7326 | /SIDSL/cgcgdata/geneseq/geneseq/NA1994.DAT:AAQ53477 + | 133.50 | 141.60 | 6.45 | 4201 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV13021 + | 135.50 | 137.61 | 10.76 | 7947 | /SIDSL/cgcgdata/geneseq/geneseq/NA1996.DAT:AAV16332 + | 133.50 | 141.60 | 6.45 | 4201 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV13020 + | 135.50 | 133.04 | 19.32 | 12438 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV99880 + | 133.50 | 141.60 | 6.45 | 4201 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV21613 + | 135.50 | 106.26 | 599.37 | 172325 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV14681 + | 133.50 | 141.09 | 6.89 | 4417 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV21612 + | 135.50 | 99.05 | 1.5e+03 | 349980 | /SIDSL/cgcgdata/geneseq/geneseq/NA1996.DAT:AAV10258 + | 133.50 | 139.16 | 8.82 | 5340 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV61724 + | 135.00 | 154.29 | 1.27 | 1453 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV71296 + | 133.50 | 139.16 | 8.82 | 5340 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV15695 + | 135.00 | 152.64 | 1.57 | 1709 | /SIDSL/cgcgdata/geneseq/geneseq/NA1996.DAT:AAV28360 + | 133.50 | 139.15 | 8.83 | 5342 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV28987 + | 135.00 | 152.64 | 2.10 | 2137 | /SIDSL/cgcgdata/geneseq/geneseq/NA1996.DAT:AAV28360 + | 133.50 | 139.15 | 8.83 | 5342 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ92234 + | 135.00 | 149.93 | 2.21 | 2229 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV14525 + | 133.50 | 139.15 | 8.83 | 5342 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AAV45910 + | 135.00 | 149.31 | 2.40 | 2369 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV24155 + | 133.50 | 139.11 | 8.88 | 5367 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV55742 + | 135.00 | 146.21 | 3.57 | 3213 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV53838 + | 133.00 | 151.47 | 1.82 | 1500 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1993.DAT:AAQ46263 + | 135.00 | 144.01 | 4.73 | 3986 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV44757 + | 133.00 | 149.61 | 2.31 | 1800 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV69717 + | 135.00 | 143.90 | 4.80 | 4031 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV64724 + | 133.00 | 148.74 | 2.58 | 1960 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV69720 + | 135.00 | 143.42 | 5.11 | 4225 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV37111 + | 133.00 | 146.38 | 3.49 | 2471 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV36149 + | 135.00 | 143.32 | 5.17 | 4255 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV78785 + | 133.00 | 145.50 | 3.91 | 2694 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV74201 + | 135.00 | 138.16 | 10.03 | 7081 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV78785 + | 133.00 | 145.50 | 3.91 | 2694 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1994.DAT:AAQ73500 + | 135.00 | 136.37 | 12.61 | 8438 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV64641 + | 133.00 | 145.43 | 3.94 | 2712 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1996.DAT:AAV31031 + | 135.00 | 135.75 | 13.65 | 8967 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV65650 + | 133.00 | 145.05 | 4.14 | 2815 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AAV87526 + | 135.00 | 135.75 | 13.65 | 8967 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV08861 + | 133.00 | 143.60 | 4.99 | 3246 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1994.DAT:AAQ70447 + | 135.00 | 133.45 | 18.33 | 11236 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV87874 + | 133.00 | 143.60 | 4.99 | 3246 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ36226 + | 135.00 | 132.53 | 20.65 | 12308 | /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AAV33215 + | 133.00 | 139.08 | 8.91 | 5061 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ36196 + | 135.00 | 132.30 | 21.24 | 12578 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV33215 + | 133.00 | 139.08 | 8.91 | 5061 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ36203 + | 135.00 | 132.09 | 21.83 | 12842 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV64850 + | 133.00 | 138.06 | 10.16 | 5595 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ36203 + | 135.00 | 131.16 | 24.61 | 14078 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV75757 + | 133.00 | 138.06 | 10.16 | 5595 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ48136 + | 135.00 | 130.80 | 25.76 | 14578 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ93709 + | 133.00 | 124.88 | 55.02 | 2038 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ36195 + | 135.00 | 130.47 | 26.89 | 15065 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ01425 + | 133.00 | 84.84 | 9.4e+03 | 1038 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV081808 + | 134.50 | 158.79 | 0.7114 | 879 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV54646 + | 132.50 | 156.40 | 0.9660 | 869 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1992.DAT:AAQ23748 + | 134.50 | 151.77 | 1.75 | 1750 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV31089 + | 132.50 | 152.63 | 1.57 | 1258 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV13836 + | 134.50 | 149.09 | 2.47 | 2277 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV16078 + | 132.50 | 148.24 | 2.75 | 1936 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV05372 + | 134.50 | 149.09 | 2.47 | 2277 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV49015 + | 132.50 | 146.48 | 3.45 | 2302 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV05372 + | 134.50 | 149.09 | 2.47 | 2277 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV11347 + | 132.50 | 142.54 | 5.71 | 3387 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV46566 + | 134.50 | 148.98 | 2.50 | 2301 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV47757 + | 132.50 | 142.25 | 5.93 | 3486 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV78912 + | 134.50 | 142.85 | 5.49 | 4200 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV67162 + | 132.50 | 142.00 | 6.12 | 3572 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV30270 + | 134.50 | 142.46 | 5.78 | 4367 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV64656 + | 132.50 | 140.33 | 7.59 | 4209 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV06552 + | 134.50 | 142.19 | 5.98 | 4481 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV93627 + | 132.50 | 137.98 | 10.26 | 5301 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAZ22348 + | 134.50 | 142.19 | 5.98 | 4481 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV93627 + | 132.50 | 137.98 | 10.26 | 5301 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1995.DAT:AAQ83525 + | 134.50 | 142.14 | 6.02 | 4507 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV4665 + | 132.50 | 136.64 | 12.18 | 6045 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV72888 + | 134.50 | 142.14 | 6.02 | 4507 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV25487 + | 132.50 | 133.20 | 18.93 | 8473 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ98282 + | 134.50 | 142.14 | 6.02 | 4507 | /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV94548 + | 132.50 | 131.91 | 22.35 | 9621 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1995.DAT:AAQ84659 + | 134.50 | 136.54 | 12.35 | 7808 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV80043 + | 132.50 | 117.17 | 147.99 | 4087 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV71718 + | 134.50 | 136.54 | 12.35 | 7808 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV31085 + | 132.00 | 151.87 | 1.73 | 1258 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV68749 + | 134.50 | 135.00 | 15.04 | 9081 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV49751 + | 132.00 | 151.34 | 1.85 | 1343 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1997.DAT:AAV79725 + | 134.50 | 134.02 | 17.05 | 9997 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV37844 + | 132.00 | 149.50 | 2.34 | 1610 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV05839 + | 134.50 | 134.02 | 17.05 | 9997 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV37844 + | 132.00 | 149.44 | 2.36 | 1610 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV45262 + | 134.00 | 155.33 | 1.11 | 1161 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAZ40490 + | 132.00 | 146.09 | 3.63 | 2249 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2001.DAT:AAV36231 + | 134.00 | 155.21 | 1.13 | 1174 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV36891 + | 132.00 | 144.57 | 4.40 | 2610 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV76856 + | 134.00 | 154.20 | 1.28 | 1297 | /SIDSL/cgcgdata/geneseq/geneseq/NA1986.DAT:AAV60473 + | 132.00 | 143.89 | 4.81 | 2790 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV14053 + | 134.00 | 153.00 | 1.49 | 1459 | /SIDSL/cgcgdata/geneseq/geneseq/NA1998.DAT:AAV32435 + | 132.00 | 142.15 | 6.01 | 3312 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV45810 + | 134.00 | 151.01 | 1.93 | 1773 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV11346 + | 132.00 | 141.93 | 6.18 | 3384 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV09015 + | 134.00 | 150.16 | 2.15 | 1929 | /SIDSL/cgcgdata/geneseq/geneseq/NA2000.DAT:AAV12454 + | 132.00 | 141.93 | 6.18 | 3384 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV09015 + | 134.00 | 148.39 | 2.70 | 2293 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV50212 + | 132.00 | 141.31 | 6.69 | 3594 |
| /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV09015 + | 134.00 | 148.39 | 2.70 | 2293 | /SIDSL/cgcgdata/geneseq/geneseq/NA1999.DAT:AAV20280 + | 132.00 | 141.15 | 6.83 | 3653 |

| | | | | | | | | | |
|--|--------|--------|---------|---------|---|--------|--------|--------|------|
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAX07348 + | 141.50 | 154.74 | 1.20 | 3089 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76340 + | 138.00 | 146.90 | 3.27 | 4341 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14691 + | 141.50 | 151.06 | 1.92 | 4433 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14683 + | 138.00 | 146.73 | 3.34 | 4414 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA93361 + | 141.50 | 150.18 | 2.15 | 4835 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA93325 + | 138.00 | 146.17 | 3.59 | 4660 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA27637 + | 141.50 | 150.18 | 2.15 | 4833 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA72751 + | 138.00 | 142.61 | 5.66 | 6008 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA21518 + | 141.50 | 148.57 | 2.64 | 5661 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA18884 + | 138.00 | 139.46 | 8.49 | 9008 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA72031 + | 141.50 | 148.49 | 2.67 | 5707 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA76857 + | 138.00 | 139.37 | 8.59 | 9089 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA70074 + | 141.50 | 148.49 | 2.67 | 5707 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA59470 + | 138.00 | 133.92 | 17.27 | 1551 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA60570 + | 141.50 | 148.49 | 2.67 | 5707 | /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA040480 + | 138.00 | 133.67 | 17.82 | 1589 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA28019 + | 141.50 | 148.20 | 2.77 | 5968 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA73805 + | 138.00 | 126.48 | 44.85 | 3220 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17482 + | 141.50 | 148.13 | 2.79 | 5910 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA22674 + | 138.00 | 121.93 | 80.39 | 5034 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA28018 + | 141.50 | 146.53 | 3.43 | 6916 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA39519 + | 138.00 | 121.93 | 80.39 | 5034 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.AAA95403 + | 141.00 | 162.78 | 0.4266 | 1320 | /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA751411 + | 138.00 | 121.54 | 84.50 | 5229 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAC84595 + | 141.00 | 157.94 | 0.7927 | 2121 | /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA047357 + | 138.00 | 121.54 | 84.50 | 5229 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA767442 + | 141.00 | 154.14 | 1.29 | 3081 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA19941 + | 138.00 | 111.69 | 298.84 | 1375 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA71719 + | 141.00 | 150.45 | 2.07 | 4424 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70096 + | 137.50 | 158.10 | 0.7769 | 1359 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14688 + | 141.00 | 150.42 | 2.08 | 4438 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70096 + | 137.50 | 154.34 | 1.26 | 1665 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA32553 + | 141.00 | 150.30 | 2.11 | 4492 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70212 + | 137.50 | 154.18 | 1.29 | 1998 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA07840 + | 141.00 | 143.28 | 5.20 | 8949 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA726543 + | 137.50 | 153.68 | 1.37 | 2097 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA28060 + | 141.00 | 141.80 | 6.29 | 10347 | /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA080908 + | 137.50 | 153.09 | 1.48 | 2223 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA26690 + | 141.00 | 127.10 | 41.43 | 43804 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA56719 + | 137.50 | 152.61 | 1.57 | 2329 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76077 + | 140.50 | 163.75 | 0.3764 | 1128 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA61632 + | 137.50 | 149.20 | 2.43 | 3257 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC51060 + | 140.50 | 160.70 | 0.5567 | 1522 | /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA91432 + | 137.50 | 148.05 | 2.82 | 3645 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA91902 + | 140.50 | 150.86 | 1.97 | 4000 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA59395 + | 137.50 | 148.05 | 2.82 | 3645 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA76173 + | 140.50 | 148.82 | 2.56 | 4886 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA140483 + | 137.50 | 147.63 | 2.98 | 3798 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA17138 + | 140.50 | 147.23 | 3.13 | 5709 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14679 + | 137.50 | 146.09 | 3.62 | 4417 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17467 + | 140.50 | 146.87 | 3.28 | 5913 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA59909 + | 137.50 | 145.14 | 4.09 | 4849 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA68385 + | 140.50 | 146.84 | 3.29 | 5934 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA287101 + | 137.50 | 144.81 | 4.27 | 5008 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA33181 + | 140.50 | 145.69 | 3.82 | 6644 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA75237 + | 137.50 | 143.16 | 5.28 | 5893 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA33182 + | 140.50 | 144.63 | 4.37 | 7372 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA59979 + | 137.50 | 143.16 | 5.28 | 5893 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA33180 + | 140.50 | 144.06 | 4.71 | 7797 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA68762 + | 137.50 | 138.65 | 9.41 | 9170 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA33184 + | 140.50 | 143.80 | 4.86 | 7996 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA52138 + | 137.50 | 134.39 | 16.25 | 1310 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA23895 + | 140.50 | 125.12 | 53.35 | 49999 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC35941 + | 137.00 | 157.85 | 0.8023 | 1310 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA31079 + | 140.00 | 162.17 | 0.4610 | 1239 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA020345 + | 137.00 | 152.83 | 1.53 | 2144 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC76302 + | 140.00 | 158.83 | 0.7072 | 1719 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA88863 + | 137.00 | 150.55 | 2.05 | 2682 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA76212 + | 140.00 | 150.49 | 2.06 | 3901 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC51192 + | 137.00 | 145.83 | 3.75 | 4263 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA54990 + | 140.00 | 149.09 | 2.47 | 4475 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAC85085 + | 137.00 | 145.35 | 3.99 | 4467 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA730092 + | 140.00 | 146.98 | 3.23 | 5502 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAC32508 + | 137.00 | 144.07 | 4.70 | 5068 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA31091 + | 139.50 | 161.57 | 0.4979 | 1236 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA59124 + | 137.00 | 142.67 | 5.62 | 5813 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.AAA083526 + | 139.50 | 155.30 | 1.11 | 2288 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA724450 + | 137.00 | 142.65 | 5.63 | 5820 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA72894 + | 139.50 | 155.30 | 1.11 | 2288 | /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.AAA029187 + | 136.50 | 151.84 | 1.73 | 2223 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA98283 + | 139.50 | 155.30 | 1.11 | 2288 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA59640 + | 136.50 | 150.44 | 2.08 | 2550 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.AAA046543 + | 139.50 | 154.88 | 1.17 | 2384 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA06553 + | 136.50 | 146.38 | 3.49 | 3798 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA09624 + | 139.50 | 154.88 | 1.17 | 2384 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17473 + | 136.50 | 141.88 | 6.22 | 5910 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA88975 + | 139.50 | 146.56 | 3.41 | 5395 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17474 + | 136.50 | 141.88 | 6.22 | 5910 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17450 + | 139.50 | 145.62 | 3.85 | 5913 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17475 + | 136.50 | 141.88 | 6.22 | 5910 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA68403 + | 139.50 | 143.09 | 5.33 | 7583 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA70206 + | 136.50 | 139.44 | 8.50 | 7503 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA93712 + | 139.50 | 127.82 | 37.78 | 39554 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC42748 + | 136.50 | 137.95 | 10.30 | 8688 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81484 + | 139.50 | 116.55 | 160.30 | 102634 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA22301 + | 136.50 | 136.99 | 11.66 | 9551 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA721609 + | 139.50 | 104.05 | 796.25 | 349980 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA98272 + | 136.50 | 136.67 | 12.13 | 9848 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA81490 + | 139.50 | 89.66 | 5.0e+03 | 1437668 | /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.AAA013593 + | 136.50 | 136.65 | 12.17 | 9871 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA72432 + | 139.00 | 162.63 | 0.4349 | 1048 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA231401 + | 136.00 | 157.33 | 0.8574 | 1219 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.AAA94725 + | 139.00 | 160.27 | 0.5879 | 1320 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA234451 + | 136.00 | 157.33 | 0.8574 | 1219 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA83134 + | 139.00 | 154.52 | 1.23 | 2322 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA64162 + | 136.00 | 154.01 | 1.31 | 1689 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA61631 + | 139.00 | 150.77 | 1.99 | 3354 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA40085 + | 136.00 | 153.04 | 1.49 | 1857 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA093345 + | 139.00 | 147.97 | 2.85 | 4416 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA34097 + | 136.00 | 152.85 | 1.52 | 1893 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA093345 + | 139.00 | 147.97 | 2.85 | 4416 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC78508 + | 136.00 | 152.84 | 1.52 | 1894 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA093345 + | 139.00 | 146.87 | 3.28 | 4920 | /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.AAA32044 + | 136.00 | 151.77 | 1.75 | 2105 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA75910 + | 139.00 | 145.03 | 4.15 | 5992 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA713955 + | 136.00 | 150.42 | 2.08 | 2403 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA17470 + | 139.00 | 145.00 | 4.17 | 5913 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA74904 + | 136.00 | 150.42 | 2.08 | 2403 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA27342 + | 138.50 | 159.34 | 0.6624 | 1360 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA68085 + | 136.00 | 150.42 | 2.08 | 2403 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA75781 + | 138.50 | 152.84 | 1.53 | 2575 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA16177 + | 136.00 | 150.42 | 2.08 | 2403 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA76844 + | 138.50 | 151.81 | 1.74 | 2848 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA13964 + | 136.00 | 150.25 | 2.13 | 2444 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC43193 + | 138.50 | 148.79 | 2.56 | 3831 | /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA74011 + | 136.00 | 150.25 | 2.13 | 2444 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAC49498 + | 138.50 | 148.33 | 2.72 | 4011 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA16190 + | 136.00 | 150.25 | 2.13 | 2444 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA24069 + | 138.50 | 147.08 | 3.19 | 4534 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA84367 + | 136.00 | 150.05 | 2.18 | 2492 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA093324 + | 138.50 | 147.08 | 3.19 | 4534 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA90855 + | 136.00 | 146.45 | 3.46 | 3546 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA7863781 + | 138.50 | 123.97 | 61.87 | 43804 | /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.AAA23429 + | 136.00 | 146.41 | 3.48 | 3563 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA42007 + | 138.00 | 165.57 | 0.2980 | 694 | /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.AAA61588 + | 136.00 | 146.31 | 3.52 | 3596 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA257866 + | 138.00 | 158.49 | 0.7391 | 1391 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14682 + | 136.00 | 144.22 | 4.61 | 4414 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.AAA73115 + | 138.00 | 150.66 | 2.02 | 3000 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14684 + | 136.00 | 144.22 | 4.61 | 4414 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA73287 + | 138.00 | 149.96 | 2.21 | 3214 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA14687 + | 136.00 | 144.22 | 4.61 | 4414 |
| /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.AAA62360 + | 138.00 | 149.96 | 2.21 | 3214 | /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA298401 + | 136.00 | 143.48 | 5.07 | 4709 |
| /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.AAA30290 + | 138.00 | 149.12 | 2.46 | 3489 | /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.AAA24028 + | 136.00 | 141.60 | 6.45 | 5709 |

| | | | | | | | | | |
|--|--------|--------|--------|--------|---|--------|--------|--------|------|
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17441 + | 153.00 | 162.51 | 0.4415 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV56447 + | 150.50 | 155.65 | 1.06 | 8532 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17442 + | 153.00 | 162.51 | 0.4415 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA93449 + | 150.50 | 155.65 | 1.06 | 8532 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT18310 + | 153.00 | 162.51 | 0.4415 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAT23891 + | 150.50 | 137.63 | 10.73 | 4999 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT18313 + | 153.00 | 162.51 | 0.4415 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAT23896 + | 150.50 | 137.63 | 10.73 | 4999 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17451 + | 153.00 | 162.51 | 0.4416 | 5915 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAV02803 + | 150.00 | 165.29 | 0.3090 | 3114 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17458 + | 153.00 | 162.51 | 0.4416 | 5915 | /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA51783 + | 150.00 | 165.15 | 0.3145 | 3156 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17460 + | 153.00 | 162.51 | 0.4416 | 5915 | /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAF25878 + | 150.00 | 163.87 | 0.3708 | 3580 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17461 + | 153.00 | 162.51 | 0.4416 | 5915 | /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC58908 + | 150.00 | 159.38 | 0.6597 | 5564 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17488 + | 153.00 | 162.51 | 0.4416 | 5915 | /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17447 + | 150.00 | 158.76 | 0.7141 | 5912 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT18332 + | 153.00 | 162.51 | 0.4416 | 5915 | /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17444 + | 150.00 | 158.76 | 0.7141 | 5912 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17436 + | 153.00 | 162.49 | 0.4425 | 5924 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68774 + | 150.00 | 154.48 | 1.24 | 8997 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT79668 + | 153.00 | 160.45 | 0.5751 | 7240 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68775 + | 150.00 | 154.48 | 1.24 | 8997 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68755 + | 153.00 | 158.18 | 0.7691 | 9044 | /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX91930 - | 150.00 | 104.38 | 763.36 | 1230 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAT25833 - | 153.00 | 130.99 | 25.15 | 130480 | /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAT68740 + | 149.50 | 154.00 | 1.32 | 8873 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC39566 + | 152.50 | 174.46 | 0.0953 | 17201 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68741 + | 149.50 | 153.84 | 1.34 | 9012 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAT28549 + | 152.50 | 133.88 | 17.36 | 92407 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68736 + | 149.50 | 153.83 | 1.34 | 9018 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAT28554 + | 152.50 | 122.98 | 70.22 | 289223 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68745 + | 149.50 | 153.83 | 1.34 | 9022 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX91242 + | 152.00 | 173.18 | 0.1123 | 1835 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT73812 + | 149.50 | 154.12 | 1.29 | 8767 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA95815 + | 152.00 | 172.21 | 0.1273 | 2019 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68740 + | 149.50 | 154.00 | 1.32 | 8873 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV17260 + | 152.00 | 167.83 | 0.2231 | 3102 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68741 + | 149.50 | 153.84 | 1.34 | 9012 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX59382 + | 152.00 | 165.76 | 0.2908 | 3800 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68736 + | 149.50 | 153.83 | 1.34 | 9018 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV22232 + | 152.00 | 163.27 | 0.4005 | 4855 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68745 + | 149.50 | 153.83 | 1.34 | 9022 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT42030 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68739 + | 149.50 | 153.82 | 1.35 | 9032 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT70073 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68744 + | 149.50 | 153.78 | 1.35 | 9066 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV60569 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68749 + | 149.50 | 153.67 | 1.37 | 9160 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV64643 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68777 + | 149.50 | 153.67 | 1.37 | 9162 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV64644 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68785 + | 149.50 | 153.67 | 1.37 | 9165 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV64649 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68778 + | 149.50 | 153.67 | 1.37 | 9165 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV64648 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68743 + | 149.50 | 153.67 | 1.37 | 9166 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV64649 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68746 + | 149.50 | 153.67 | 1.37 | 9166 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV64642 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68738 + | 149.50 | 153.67 | 1.37 | 9166 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV11737 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68769 + | 149.50 | 153.66 | 1.37 | 9167 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC60795 + | 152.00 | 161.61 | 0.4952 | 5711 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68742 + | 149.50 | 153.66 | 1.37 | 9167 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17479 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68780 + | 149.50 | 153.66 | 1.37 | 9168 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17473 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68776 + | 149.50 | 153.66 | 1.37 | 9168 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17484 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT73808 + | 149.50 | 153.66 | 1.37 | 9169 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17493 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT73811 + | 149.50 | 153.66 | 1.37 | 9169 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17502 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68763 + | 149.50 | 153.66 | 1.37 | 9169 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17508 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68766 + | 149.50 | 153.66 | 1.37 | 9169 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17539 + | 152.00 | 161.26 | 0.5183 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68734 + | 149.50 | 153.66 | 1.37 | 9169 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX35273 - | 152.00 | 160.73 | 0.5542 | 6225 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68731 + | 149.50 | 153.66 | 1.37 | 9169 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT69707 + | 152.00 | 160.73 | 0.5542 | 6225 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT73497 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV21209 + | 152.00 | 156.98 | 0.8966 | 8996 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT73497 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX30256 + | 152.00 | 155.42 | 1.10 | 10485 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT73810 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX30257 + | 152.00 | 155.42 | 1.10 | 10485 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68781 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX30258 + | 152.00 | 155.42 | 1.10 | 10485 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68782 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX30259 + | 152.00 | 155.42 | 1.10 | 10485 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68784 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX30255 + | 152.00 | 155.42 | 1.10 | 10485 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68765 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT69707 + | 152.00 | 154.58 | 1.22 | 11385 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68768 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV21209 + | 152.00 | 166.65 | 0.2594 | 3275 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68773 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAC99113 + | 151.50 | 166.27 | 0.2724 | 3399 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68737 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC99113 + | 151.50 | 165.44 | 0.3032 | 3890 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68733 + | 149.50 | 153.66 | 1.37 | 9171 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC930849 + | 151.50 | 163.36 | 0.3958 | 4325 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68748 + | 149.50 | 153.66 | 1.37 | 9172 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA96975 + | 151.50 | 160.59 | 0.5648 | 5940 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68732 + | 149.50 | 153.66 | 1.37 | 9172 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAT70105 + | 151.50 | 160.59 | 0.5648 | 5940 | /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV30198 + | 149.50 | 151.02 | 1.93 | 1188 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAZ24616 + | 151.50 | 156.41 | 0.9647 | 8948 | /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV30198 + | 149.50 | 150.28 | 2.12 | 1284 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC65855 + | 151.50 | 156.41 | 0.9647 | 8948 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68784 + | 149.50 | 150.28 | 2.12 | 1284 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA81498 + | 151.50 | 155.74 | 1.0647 | 9560 | /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAZ31081 + | 149.00 | 173.06 | 0.1140 | 1377 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAF57301 + | 151.50 | 154.66 | 1.21 | 10628 | /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAF25866 + | 149.00 | 162.62 | 0.4353 | 3580 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT.AAT73436 + | 151.50 | 139.54 | 8.40 | 46899 | /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT.AAQ06328 + | 149.00 | 162.11 | 0.4646 | 3763 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX20248 + | 151.50 | 103.82 | 40.5 | 910715 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT80072 + | 149.00 | 162.11 | 0.4646 | 3763 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT49317 + | 151.00 | 173.22 | 0.1118 | 1617 | /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT.AAQ04786 + | 149.00 | 160.95 | 0.5395 | 4219 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAD00601 + | 151.00 | 165.66 | 0.2947 | 3395 | /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAV59103 + | 149.00 | 158.56 | 0.7322 | 5330 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17463 + | 151.00 | 160.01 | 0.6084 | 5913 | /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT.AAQ030251 + | 149.00 | 158.29 | 0.7581 | 5474 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17443 + | 151.00 | 160.01 | 0.6085 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT34231 + | 149.00 | 157.06 | 0.8874 | 6175 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17478 + | 151.00 | 160.01 | 0.6085 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT68779 + | 149.00 | 153.02 | 1.49 | 9180 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT18311 + | 151.00 | 160.01 | 0.6085 | 5914 | /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT.AAQ27234 + | 149.00 | 152.56 | 1.58 | 9606 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAT75529 + | 150.50 | 170.64 | 0.1556 | 1959 | /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT95538 + | 149.00 | 152.56 | 1.58 | 9606 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.AAF45088 + | 150.50 | 168.39 | 0.2077 | 2443 | /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.AAT96153 + | 149.00 | 152.56 | 1.58 | 9606 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.AAC207192 + | 150.50 | 165.40 | 0.3046 | 3275 | /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT.AAT9612297 + | 149.00 | 151.72 | 1.76 | 1043 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC79145 + | 150.50 | 165.40 | 0.3046 | 3275 | /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAC58912 + | 148.50 | 157.48 | 0.8409 | 5573 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.AAT17477 + | 150.50 | 159.38 | 0.6592 | 5913 | /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA58986 + | 148.50 | 155.81 | 1.04 | 6568 |
| /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.AAA33450 + | 150.50 | 156.02 | 1.02 | 8229 | /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT.AAA92896 + | 148.00 | 169.09 | 0.1897 | 1677 |

THIS PAGE BLANK (USPTO)


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unmayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL; AC006201; AAD20129.1; -.
SQ SEQUENCE 507 AA; 57427 MW; B073165433A764CF CRC64;

Query Match 0.6%; Score 8; DB 10; Length 507;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332
Db 153 EEEEEEE 160
|||||
153 EEEEEEE 160

RESULT 49
Q9VS02 PRELIMINARY; PRT; 508 AA.
AC Q9VS02;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG9953 PROTEIN.
GN CG9953.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC 1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR EMBL; AE003560; AAF50628.1; -.
DR FlyBase; FBgn0035726; CG9953.
DR InterPro; IPR000073; -.
DR InterPro; IPR000379; -.
DR Pfam; PF00561; abhydrolase_1.
SQ SEQUENCE 508 AA; 56861 MW; 4AF644136F22F00B CRC64;

Query Match 0.6%; Score 8; DB 5; Length 508;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 KPXTNNVL 1131
Db 431 KPXTNNVL 438
|||||
431 KPXTNNVL 438

RESULT 50
Q75239 PRELIMINARY; PRT; 510 AA.
ID Q75239;
AC Q75239;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE F23149_1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 2.2 Mb region in 19q13.1 containing the RYR
RT gene.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005239; AAC25503.1; -.
SQ SEQUENCE 510 AA; 57466 MW; CACE73EDC7290CE8 CRC64;

Query Match 0.6%; Score 8; DB 4; Length 510;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEPEEEEE 1331
Db 355 PEPEEEEE 362
|||||
355 PEPEEEEE 362

```

Search completed: September 25, 2001, 10:07:44
Job time: 175 sec

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

DB 98 PEEEEEE 105

RESULT 45

ID O43818 PRELIMINARY; PRT; 475 AA.

AC O43818;

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE U3 SNORNP ASSOCIATED 55 KDA PROTEIN.

GN U3-55K.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98078705; PubMed=9418896;

RA Pluk H., Soffner J., Luhrmann R., Van Venrooij W.J.;

RT "CDNA cloning and characterization of the human U3 small nucleolar

ribonucleoprotein complex-associated 55-kilodalton protein.";

RL Mol. Cell. Biol. 18:488-498(1998).

DR EMBL; AJ001340; CAA04687.1; -.

DR InterPro; IPR001680; -.

DR Pfam; PF00400; WD40; 6.

DR PRINTS; PR00320; GPROTEINBRPT.

DR PROSITE; PS00678; WD-REPEATS; UNKNOWN_1.

DR SMART; SM00320; WD40; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 475 AA; 51840 MW; 3CD19F66EA75B627 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 4; Length 475;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331

DB 63 PEEEEEE 70

RESULT 46

Q39482

IL Q39482 PRELIMINARY; PRT; 501 AA.

AC Q39482;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE LEGUMIN (FRAGMENT).

GN LEG1.

OS Calocedrus decurrens.

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Coniferopsida; Coniferales; Cupressaceae; Calocedrus.

OX NCBI_TaxID=13387;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=IMMATURE SEED;

RA Haeger K.P., Dank N.;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; X95539; CAA64786.1; -.

DR Mendel; 7560; Calde; 1189; 7560.

DR InterPro; IPR000459; -.

DR InterPro; IPR001596; -.

DR Pfam; PF00190; Seedstore_11s; 3.

DR PRINTS; PR00439; 11SGLOBULIN.

DR PROSITE; PS00387; PPASE; UNKNOWN_1.

KW Seed storage protein.

FT NON_TER 1

SQ SEQUENCE 501 AA; 56975 MW; 7877D42C60E9D6E5 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 10; Length 501;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332

DB 140 EEEEEEE 147

RESULT 47

ID P97770

AC P97770 PRELIMINARY; PRT; 505 AA.

DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE HYPOTHETICAL 56.4 KDA PROTEIN (ROSA26AS).

GN GTROSA26AS.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91357468; PubMed=1653172;

RA Friedrich G.A., Soriano P.;

RT "Promoter traps in embryonic stem cells: a genetic screen to identify

and mutate developmental genes in mice.";

RL Genes Dev. 5:1513-1523(1991).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97268648; PubMed=9108056;

RA Zambrowicz B.P., Inamoto A., Fiering S., Herzenberg L.A., Kerr W.G.,

RA Soriano P.;

RT "Disruption of overlapping transcripts in the ROSA beta geo 26 gene

trap strain leads to widespread expression of beta-galactosidase in

mouse embryos and hematopoietic cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 94:3789-3794(1997).

DR EMBL; U83176; AAC60384.1; -.

DR MGD; MGI:1277973; Gtrosa26as.

DR InterPro; IPR000241; -.

DR Pfam; PF01170; UPF0020; 1.

DR PROSITE; PS01261; UPF0020; UNKNOWN_1.

KW Hypothetical protein.

SQ SEQUENCE 505 AA; 56431 MW; 6DA0F4C5B768C541 CRC64;

Query Match

Best Local Similarity 0.6%; Score 8; DB 11; Length 505;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 LLKRTAQA 228

DB 483 LLKRTAQA 490

RESULT 48

Q9SL49

ID Q9SL49 PRELIMINARY; PRT; 507 AA.

AC Q9SL49;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE AT2G17970 PROTEIN.

GN AT2G17970.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

DE CHROMOGRANIN A PRECURSOR.
GN CHGA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ENGLAND DEACONNESS HOSPITAL;
RX MEDLINE=90007662; PubMed=2793216;
RA Farmer R.J., Koop A.H., Handa M.T., O'Connor D.T.;
RT "Molecular cloning of chromogranin A from rat pheochromocytoma
cells";
RL Hypertension 14:435-444(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ENGLAND DEACONNESS HOSPITAL;
RA Farmer R.J., Koop A.H., Handa M.T., O'Connor D.T.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF145445; AAD40652.1; -.
DR HSP; P05059; 1CFK.
DR InterPro; IPR001819; -.
DR InterPro; IPR001990; -.
DR Pfam; PF01271; Granin; 1.
DR PRINTS; PR00659; CHROMOGRANIN.
DR PROSITE; PS00422; GRANINS_1; 1.
DR PROSITE; PS00423; GRANINS_2; 1.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 381 401 CATESTATIN.
SQ SEQUENCE 462 AA; 51512 MW; DA3F253DEC3F6282 CRC64;

Query Match 0.6%; Score 8; DB 11; Length 462;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 340 EEEEEER 347

RESULT 42
Q9HEL6 PRELIMINARY; PRT; 466 AA.
AC Q9HEL6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN 12F11.100.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakates G., Mewes H.W., Mannhaupt G.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL451017; CAC18219.1; -.
SQ SEQUENCE 466 AA; 51449 MW; ED901EFCEC7ADCBD CRC64;

Query Match 0.6%; Score 8; DB 3; Length 466;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1323 APEEEEE 1330

DB 448 APEEEEE 455
RESULT 43
O57853 PRELIMINARY; PRT; 468 AA.
ID O57853;
AC O57853;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE 468AA LONG HYPOTHETICAL REPLICATION FACTOR C SUBUNIT.
GN PH0113.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3";
RL DNA Res. 5:55-76(1998).
DR EMBL; AF000001; BAA29182.1; -.
DR InterPro; IPR000862; -.
DR InterPro; IPR001939; -.
DR InterPro; IPR003593; -.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
SQ SEQUENCE 468 AA; 53913 MW; B85739AD2760B5AF CRC64;

Query Match 0.6%; Score 8; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 436 EEEEEER 443

RESULT 44
Q9QYK4 PRELIMINARY; PRT; 470 AA.
ID Q9QYK4;
AC Q9QYK4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HEPARAN SULFATE 6-SULFOTRANSFERASE 3.
GN HS6ST3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Habuchi H.;
RT "Mouse, heparan sulfate 6-O-sulfotransferase";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB024567; BAA89249.1; -.
DR MGD; MGI:1354960; Hs6st3.
KW Transferase.
SQ SEQUENCE 470 AA; 55071 MW; 2D5EB0A82B895A66 CRC64;

Query Match 0.6%; Score 8; DB 11; Length 470;
Best Local Similarity 100.0%; Pred. No. 40;

DR Pfam; PF00621; RhoGEF; 1.
 FT NON_TER 425 425
 SQ SEQUENCE 425 AA; 49871 MW; C718616905736E7F CRC64;

Query Match 0.6%; Score 8; DB 4; Length 425;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
 DQ 11111111
 57 PEEEEEE 64

RESULT 39
 O34631 PRELIMINARY; PRT; 442 AA.

AC O34631;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE YVRA PROTEIN.
 GN YVRA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,
 Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 Presecan E., Pujic P., Furnelle B., Rapoport G., Rey M., Reynolds S.,
 Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 Viari A., Wambutt R., Wedler H., Wedler H., Weitzneger T.,
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 *The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis*;
 Nature 390:249-256(1997).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 Submitted (NOV-1997) to the EMBL/GenBank/DBDJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Wipat A., Brignell C.S., Guy J.B., Rose M., Emerson P.T.,
 RA Harwood C.R.;
 Submitted (FEB-1998) to the EMBL/GenBank/DBDJ databases.

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Rose M., Entian K.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBDJ databases.
 DR EMBL; Z99120; CAB15306.1; -;
 DR EMBL; Z99121; CAB15321.1; -;
 DR EMBL; AJ223978; CAAL1737.1; -;
 DR InterPro; IPR001687; -;
 DR InterPro; IPR003439; -;
 DR InterPro; IPR003593; -;
 DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 SQ SEQUENCE 442 AA; 48631 MW; 8020C6870E43C7AE CRC64;

Query Match 0.6%; Score 8; DB 2; Length 442;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 KDLLDLIK 514
 DQ 11111111
 174 KDLLDLIK 181

RESULT 40
 Q9WVR7 PRELIMINARY; PRT; 450 AA.

AC Q9WVR7;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE CA/CALMODULIN-DEPENDENT PROTEIN KINASE PHOSPHATASE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=99279559; PubMed=10348902;
 RA Kitani T., Ishida A., Okuno S., Takeuchi M., Kameshita I.,
 RA Fujisawa H.;
 RT "Molecular Cloning of Ca2+/Calmodulin-dependent protein kinase
 phosphatase";
 RL J. Biochem. 125:1022-1028(1999).
 DR EMBL; AB023634; BAA82477.1; -;
 DR InterPro; IPR000222; -;
 DR InterPro; IPR001932; -;
 DR Pfam; PF00481; PP2C; 1.
 DR PROSITE; PS01032; PP2C; 1.
 DR SMART; SM00331; PP2C_SIG; 1.
 KW Kinase.
 SQ SEQUENCE 450 AA; 49165 MW; 67626542F32B2FD0 CRC64;

Query Match 0.6%; Score 8; DB 11; Length 450;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
 DQ 11111111
 103 EEEEEER 110

RESULT 41
 Q9R1B7 PRELIMINARY; PRT; 462 AA.

AC Q9R1B7;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

KW Hypothetical protein.
SQ SEQUENCE 379 AA; 40158 MW; CCE6BF58679A8142 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 379;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 99 EEEEEER 106

RESULT 35
Q9GL18 PRELIMINARY; PRT; 388 AA.
AC Q9GL18;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ALPHA ADRENERGIC RECEPTOR 2B (FRAGMENT).
GN AAR2B.
OS Diceros bicornis (Black rhinoceros).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Diceros.
OX NCBI_TaxID=9805;
RN [1]
RP SEQUENCE FROM N.A.
RA Madsen O., Kao D., De Bry R.W., Douady C., Stanhope M.J.,
de Jong W.W., Springer M.S.;
RT "Molecules untangle the basal divergences of placental mammals.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ251184; CAC16686.1;
KW Receptor.
FT NON_TER 1 1
FT NON_TER 388 388
SQ SEQUENCE 388 AA; 42546 MW; 14EB08C6FB4C5371 CRC64;

Query Match 0.6%; Score 8; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEER 1331
DB 279 PEEEEER 286

RESULT 36
Q9GK25 PRELIMINARY; PRT; 389 AA.
AC Q9GK25;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ALPHA ADRENERGIC RECEPTOR 2B (FRAGMENT).
GN AAR2B.
OS Tupala tana (large tree shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Scandentia; Tupaiidae; Tupaiia.
OX NCBI_TaxID=70687;
RN [1]
RP SEQUENCE FROM N.A.
RA Madsen O., Kao D., De Bry R.W., Douady C., Stanhope M.J.,
de Jong W.W., Springer M.S.;
RT "Molecules untangle the basal divergences of placental mammals.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ251187; CAC16698.1;
KW Receptor.
FT NON_TER 1 1
FT NON_TER 389 389
SQ SEQUENCE 389 AA; 42631 MW; A9DC6F864487A2BE CRC64;

Query Match 0.6%; Score 8; DB 6; Length 389;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEER 1331
DB 281 PEEEEER 288

RESULT 37
Q9LTR6 PRELIMINARY; PRT; 395 AA.
AC Q9LTR6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE GB|AAF36750.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty P1 and TAC
clones.";
RT DNA Res. 7:131-135(2000).
DR EMBL; AB024036; BAB02807.1;
SQ SEQUENCE 395 AA; 45347 MW; C4419188AD62622D CRC64;

Query Match 0.6%; Score 8; DB 10; Length 395;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEER 1331
DB 79 PEEEEER 86

RESULT 38
Q9NQD6 PRELIMINARY; PRT; 425 AA.
AC Q9NQD6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR (FRAGMENT).
GN GEF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Rodrigues N.R., Theodosiou A.M., Nesbit M.A., Tandle A.T.,
Saranath D., Davies K.E.;
RT "A novel member of the Dbl family of oncogenes expressed predominantly
in the caudate nucleus.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238899; CAC00686.1;
DR InterPro; IPR000219;


```

RESULT 31
Q9YFQ6 PRELIMINARY; PRT; 341 AA.
ID Q9YFQ6;
AC Q9YFQ6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 36.4 KDA PROTEIN APE0193.
GN APE0193.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310139; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
RL EMBL; AP000058; BAA79105.1; -.
KW Hypothetical protein.
SQ SEQUENCE 341 AA; 36378 MW; ACD9D8D78A438B85 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 981 EKLLSLP 988
Db 325 EKLLSLP 332

RESULT 32
O96263 PRELIMINARY; PRT; 351 AA.
ID O96263;
AC O96263;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RNA METHYLASE (SPOU FAMILY) (OO, TP).
GN PF0855C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RF SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
RL EMBL; AE001421; AAC71960.1; -.
DR InterPro; IPR001537; -.
DR Pfam; PF00588; Spou_methylase; 1.
DR ProDom; PD001243; -.
KW Methyltransferase.
SQ SEQUENCE 351 AA; 41783 MW; 18EA09F50EFA4396 CRC64;

Query Match 0.6%; Score 8; DB 5; Length 351;
Best Local Similarity 100.0%; Pred. No. 31;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1270 LKEDILEN 1277
Db 305 LKEDILEN 312

RESULT 33
Q9P837 PRELIMINARY; PRT; 368 AA.
ID Q9P837;
AC Q9P837;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ACTIN BINDING PROTEIN (FRAGMENT).
GN ABP1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RF SEQUENCE FROM N.A.
RA De Backer M.D., Logghe M., Viaene J., Loonen I., Vandoninck S.,
RA de Hoogt R., Nelissen B., Dewaele S., Simons F., Verhasselt P.,
RA Contreras R., Luytens W.H.M.L.;
RT "A novel method for systematic identification of genes required for
RT growth of Candida albicans."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ390503; CAB77642.1; -.
DR InterPro; IPR000108; -.
DR InterPro; IPR001452; -.
DR InterPro; IPR002965; -.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 2.
FT NON_TER 1
SQ SEQUENCE 368 AA; 40241 MW; E1FC753CF1DC197F CRC64;

Query Match 0.6%; Score 8; DB 3; Length 368;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEE 1331
Db 176 PEEEEE 183

RESULT 34
Q9XE25 PRELIMINARY; PRT; 379 AA.
ID Q9XE25;
AC Q9XE25;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 40.2 KDA PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
RT clone:PO026F07."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AP000364; BAA81779.1; -.
DR InterPro; IPR000847; -.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.

```



```

QY 1325 EEEEEER 1332
DB 114 EEEEEER 121

RESULT 28
O44016 PRELIMINARY; PRT; 325 AA.
AC O44016;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 38.4 KDA PROTEIN (G5 ORF).
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
OA NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=98198836; PubMed=9539429;
RA Rieben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.;
RA "Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to
RA the Ddpl and Ddp2 plasmid families.";
RA Genetics 148:1117-1125(1998).
DR EMBL; U00796; AAC18634.1; -
KW Hypothetical protein.
SQ SEQUENCE 325 AA; 38447 MW; 69A43D0C632058A6 CRC64;

Query Match 0.6%; Score 8; DB 5; Length 325;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
DB 159 PEEEEEE 166

RESULT 29
Q9FM77 PRELIMINARY; PRT; 329 AA.
AC Q9FM77;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MDF20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OA NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RA "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RA Sequence features of the regions of 1,456,315 bp covered by nineteen
RA physically assigned pl and TAC clones.";
RA DNA Res. 5:41-54(1998).
DR EMBL; AB09050; BAB09228.1; -
SQ SEQUENCE 329 AA; 36055 MW; A5AA93B33FD7A44B CRC64;

Query Match 0.6%; Score 8; DB 10; Length 329;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
DB 114 EEEEEER 121

```

```

DB 249 PEEEEEE 256

RESULT 30
Q9VD43 PRELIMINARY; PRT; 331 AA.
AC Q9VD43;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CG7087 PROTEIN.
GN CG7087.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OA NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Amanatides P.G., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003738; AAF55960.1; -
DR FlyBase; FBgn0038935; CG7087.
SQ SEQUENCE 331 AA; 36524 MW; 3A805B2FA07468F6 CRC64;

```

```

Query Match 0.6%; Score 8; DB 5; Length 331;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 287 EEEEEER 294

```


GN PAC3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=WS;
RX MEDLINE=95003710; PubMed=7919990;
RA Reiter R.S., Comber S.A., Bourett T.M., Bartley G.E., Scolnik P.A.;
RT "Control of leaf and chloroplast development by the Arabidopsis gene
pale cress.";
RI Plant Cell 6:1253-1264(1994).
RL EMBL; L35241; AAA21761.1; -
DR Mendel; 17382; Arath;2781;17382.
SQ SEQUENCE 310 AA; 36145 MW; 47567975BEFFFC41 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 310;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 80 EEEEEER 87

RESULT 25
ID Q39089 PRELIMINARY; PRT; 313 AA.
AC Q39089;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PALE CRESS GENE.
GN PALE CRESS OR AT2G48120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=96323027; PubMed=8709959;
RA Greveling C., Suter-Crazzolara C., Menges A., von Kempner E.,
RA Masterson R., Schell J., Reiss B.;
RT "Characterisation of a new allele of pale cress and its role in
greening in Arabidopsis thaliana.";
RL Mol. Gen. Genet. 251:532-541(1996).
RN [2]
RE SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL; X96480; CAA65334.1; -
DR EMBL; X96482; CAA65337.1; -
DR EMBL; AC006072; AAD48512.1; -
DR Mendel; 17381; Arath;2781;17381.
SQ SEQUENCE 313 AA; 36611 MW; E3711AA2A4E975BE CRC64;

Query Match 0.6%; Score 8; DB 10; Length 313;
Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 80 EEEEEER 87

RESULT 26
ID Q50887 PRELIMINARY; PRT; 316 AA.
AC Q50887;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CARD GENE AND OPEN READING FRAMES.
GN CARD.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=DK1050;
RX MEDLINE=96293442; PubMed=8692912;
RA Nicolas F.J., Cayuela M.L., Martinez-Argudo I.M., Ruiz-Vazquez R.M.,
RA Murillo F.J.;
RT "High mobility group I(Y)-like DNA-binding domains on a bacterial
transcription factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6881-6885(1996).
DR EMBL; 256280; CAA91224.1; -
DR InterPro; IPR000637; -
DR Pfam; PF02178; AT_hook; 4.
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 316 AA; 34029 MW; B4277B36B2A60858 CRC64;

Query Match 0.6%; Score 8; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1289 KGRPPKP 1296
DB 257 KGRPPKP 264

RESULT 27
ID Q9LI19 PRELIMINARY; PRT; 316 AA.
AC Q9LI19;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0708G02.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001539; BAA92919.1; -
SQ SEQUENCE 316 AA; 33263 MW; 7D2BFA7A90C9C8B0 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 316;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE DOF ZINC FINGER PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 OC Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YUKTHIKARI; TISSUE=ALEURONE LAYER;
 RX MEDLINE=9418329; PubMed=10490396;
 RA Washio K.;
 RT "CDNAs encoding Dof-proteins that are present in germinated aleurone
 cells (Accession Nos. AB028129, AB028130, AB028131, AB028132, and
 AB028133). (PGR99-107).";
 RL Plant Physiol. 120:1205-1205(1999).
 DR EMBL; AB028133; BAA78576.1; -.
 DR InterPro: IPR000345; -.
 DR PROSITE: PS00190; CYNCHROME_C; UNKNOWN1.
 SQ SEQUENCE 235 AA; 25410 MW; D5B383072B30F540 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 235;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
 DB 53 PEEEEEE 60

RESULT 21
 O95472 PRELIMINARY; PRT; 251 AA.
 ID O95472;
 AC O95472;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE SLOW SKELETAL MUSCLE TROPONIN T.
 GN TNNT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Barton P.J.R., Cullen M.E., Townsend P.J., Brand N.J., Mullen A.J.,
 RA Norman D.A.M., Bhavsar P.K., Yacoub M.H.;
 RT "Close physical linkage of human troponin genes: organization,
 RT sequence and expression of the locus encoding cardiac troponin I and
 RT slow skeletal troponin T.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AJ011712; CAA09750.1; -.
 DR EMBL; AJ011713; CAA09750.1; JOINED.
 DR InterPro: IPR001978; -.
 DR Pfam: PF00992; Troponin; 1.
 SQ SEQUENCE 251 AA; 30096 MW; DA9E80D5F8B5DE19 CRC64;

Query Match 0.6%; Score 8; DB 4; Length 251;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEEE 1332
 DB 21 EEEEEEE 28

RESULT 22
 O22924 PRELIMINARY; PRT; 294 AA.
 ID O22924

AC O22924;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
 DE T09D09.9 PROTEIN.
 GN T09D09.9
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
 RA Venter J.C.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC002338; AAC16933.1; -.
 SQ SEQUENCE 294 AA; 34018 MW; 06217721EEEF2402 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 294;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
 DB 226 PEEEEEE 233

RESULT 23
 O9M4F1 PRELIMINARY; PRT; 299 AA.
 ID O9M4F1;
 AC O9M4F1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SIAH2 PROTEIN.
 GN SIAH2.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. W1;
 RA Cui Y., Brugiere N., Jackman L., Bi Y.M., Rothstein S.J.;
 RT "Structural and transcriptional comparative analysis of the S-locus
 RT regions in two self-incompatible Brassica napus lines.";
 RL Plant Cell 11:2217-2231(2000).
 DR EMBL; AJ245479; CAB89184.1; -.
 SQ SEQUENCE 299 AA; 33686 MW; D5BEC8646E3111DD CRC64;

Query Match 0.6%; Score 8; DB 10; Length 299;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 PEEEEEE 1331
 DB 25 PEEEEEE 32

RESULT 24
 Q39175 PRELIMINARY; PRT; 310 AA.
 ID Q39175
 AC Q39175;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE PALE CRESS PAC3.

[illegible]


```
DR PRINTS; PR000929; ATHOOK.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 99 AA; 10472 MW; 5BA44A226B8A6990 CRC64;

Query Match 0.6%; Score 8; DB 5; Length 99;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 KGRGPSK 1380
DB 55 KGRGPSK 62

RESULT 13
Q23794 PRELIMINARY; PRT; 99 AA.
ID Q23794
AC Q23794;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN I/Y.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=95105193; PubMed=7806532;
RA Claus P., Schulze E., Wisniewski J.R.;
RT "Insect proteins homologous to mammalian high mobility group proteins
RT I/Y (HMG I/Y). Characterization and binding to linear and four-way
RT junction DNA."; 269:33042-33048(1994).
RL J. Biol. Chem. 269:33042-33048(1994).
DR EMBL; Z36898; CAA85365.1; -
DR InterPro; IPR000637; -
DR Pfam; PF02178; AT_hook; 3.
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 99 AA; 10502 MW; 1AA4A226B8A698C CRC64;

Query Match 0.6%; Score 8; DB 5; Length 99;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1373 KGRGPSK 1380
DB 55 KGRGPSK 62

RESULT 14
Q96475 PRELIMINARY; PRT; 131 AA.
ID Q96475
AC Q96475;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CONGLUTIN ALPHA (CONALPHA) (FRAGMENT).
GN CONALPHA.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Lupinus.
OX NCBI_TaxID=3871;
RN [1]
RS SEQUENCE FROM N.A.
RX STRAIN=UNICROP.
RA MEDLINE=97390686; PubMed=9247543;
RA Ilgoutz S.C., Knittel N., Lin J.M., Sterle S., Gayler K.R.;
RT "Transcription of genes for conglutin gamma and a leginsulin-like
RT protein in narrow-leaved lupin.";

Plant Mol. Biol. 34:613-627(1997).
-!- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER IIS SEED STORAGE PROTEINS (GLOBULINS).
DR EMBL; U74384; AAC49787.1; -
DR Mengel; 24412; Lupan; 1189; 24412.
DR InterPro; IPR000459; -
DR Pfam; PF00190; Seedstore_11s; 1.
DR PRINTS; PR00439; IISGLOBULIN.
DR PROSITE; PS00305; IIS_SEED_STORAGE; 1.
KW Multigene family.
FT NON_TER 1
FT NON_TER 131
SQ SEQUENCE 131 AA; 15532 MW; 75BF5C9609AD9F09 CRC64;

Query Match 0.6%; Score 8; DB 10; Length 131;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 8 EEEEEER 15

RESULT 15
Q84290 PRELIMINARY; PRT; 147 AA.
ID Q84290
AC Q84290;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TYPE 4 E4 PROTEIN.
GN E4.
OS Human papillomavirus.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10566;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=89370332; PubMed=2549722;
RA Doorbar J., Coneron I., Gallimore P.H.;
RT "Sequence divergence yet conserved physical characteristics among the
RT E4 proteins of cutaneous human papillomaviruses.";
RL Virology 172:51-62(1989).
DR EMBL; M28744; AAA47015.1; -
SQ SEQUENCE 147 AA; 16854 MW; A238A5DCEEC8C9FD CRC64;

Query Match 0.6%; Score 8; DB 14; Length 147;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1325 EEEEEER 1332
DB 102 EEEEEER 109

RESULT 16
Q9LV56 PRELIMINARY; PRT; 149 AA.
ID Q9LV56
AC Q9LV56;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE GB|AAD25142.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
```



```
DR InterPro: IPR000637; -.
DR Pfam: PF02178; AT_hook; 3.
DR PRINTS: PRO0929; ATHOOK.
DR SMART: SM00384; AT_hook; 1.
DR SC SEQUENCE 837 AA; 90444 MW; 140C484C027036BF CRC64;

Query Match          0.6%; Score 9; DB 3; Length 837;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1287 KGKGRPPK 1295
Db 695 KGKGRPPK 703
|||||

RESULT 9
QSRIN5 PRELIMINARY; PRT; 59 AA.
AC QSRIN5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE GLUCOSIDASE II BETA-SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99150222; PubMed=10024665;
RA Arendt C.W., Dawicki W., Ostergaard H.L.;
RT "Alternative splicing of transcripts encoding the alpha- and beta-
RT subunits of mouse glucosidase II in T lymphocytes.";
RL Glycobiology 9:277-283(1999).
DR EMBL: AF066061; AAD43365.1; -.
FT NON_TER 1
FT NON_TER 59
FT SEQUENCE 59 AA; 6878 MW; C64B99BD0D94826 CRC64;

Query Match          0.6%; Score 8; DB 11; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db 39 PEEEEEE 46
|||||

RESULT 10
QSRVP2 PRELIMINARY; PRT; 66 AA.
AC QSRVP2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GLUCOSIDASE II BETA-SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99150222; PubMed=10024665;
RA Arendt C.W., Dawicki W., Ostergaard H.L.;
RT "Alternative splicing of transcripts encoding the alpha- and beta-
RT subunits of mouse glucosidase II in T lymphocytes.";
RL Glycobiology 9:277-283(1999).
DR EMBL: AF066061; AAD43364.1; -.
FT NON_TER 1
FT NON_TER 66
FT SEQUENCE 66 AA; 7645 MW; 3B655092C1355B3B CRC64;

Query Match          0.6%; Score 9; DB 3; Length 837;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1287 KGKGRPPK 1295
Db 695 KGKGRPPK 703
|||||

RESULT 9
QSRIN5 PRELIMINARY; PRT; 59 AA.
AC QSRIN5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE GLUCOSIDASE II BETA-SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99150222; PubMed=10024665;
RA Arendt C.W., Dawicki W., Ostergaard H.L.;
RT "Alternative splicing of transcripts encoding the alpha- and beta-
RT subunits of mouse glucosidase II in T lymphocytes.";
RL Glycobiology 9:277-283(1999).
DR EMBL: AF066061; AAD43365.1; -.
FT NON_TER 1
FT NON_TER 59
FT SEQUENCE 59 AA; 6878 MW; C64B99BD0D94826 CRC64;

Query Match          0.6%; Score 8; DB 11; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db 39 PEEEEEE 46
|||||

RESULT 10
QSRVP2 PRELIMINARY; PRT; 66 AA.
AC QSRVP2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GLUCOSIDASE II BETA-SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99150222; PubMed=10024665;
RA Arendt C.W., Dawicki W., Ostergaard H.L.;
RT "Alternative splicing of transcripts encoding the alpha- and beta-
RT subunits of mouse glucosidase II in T lymphocytes.";
RL Glycobiology 9:277-283(1999).
DR EMBL: AF066061; AAD43364.1; -.
FT NON_TER 1
FT NON_TER 66
FT SEQUENCE 66 AA; 7645 MW; 3B655092C1355B3B CRC64;

Query Match          0.6%; Score 8; DB 11; Length 66;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1324 PEEEEEE 1331
Db 39 PEEEEEE 46
|||||

RESULT 11
QSRM3G5 PRELIMINARY; PRT; 81 AA.
AC QSRM3G5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 8.9 KDA PROTEIN.
GN T14K23_20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nyakatura G., Fartmann B., Dauner D., Sterr W., Holland R.,
RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RA Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132909; CAB87723.1; -.
KW Hypothetical protein.
SQ SEQUENCE 81 AA; 8929 MW; 1B2371B2F081ABEE CRC64;

Query Match          0.6%; Score 8; DB 10; Length 81;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 PEEEEEE 1332
Db 19 PEEEEEE 26
|||||

RESULT 12
QSR3793 PRELIMINARY; PRT; 99 AA.
AC QSR3793;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN I/Y.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-95105193; PubMed=7806532;
RA Claus P., Schulze E., Wisniewski J.R.;
RT "Insect proteins homologous to mammalian high mobility group proteins
RT I/Y (HMG I/Y): Characterization and binding to linear and four-way
RT junction DNA.";
RL J. Biol. Chem. 269:33042-33048(1994).
DR EMBL: Z36897; CAA85364.1; -.
DR InterPro: IPR000637; -.
DR Pfam: PF02178; AT_hook; 3.
```


Proc. Am. Assoc. Cancer Res. 37:223-223(1996).

RE
RN
RC
RX
RA
RT
RT
RT
RL
DR
DR
DR
SQ

[12]
SEQUENCE FROM N.A.
TISSUE=PROSTATE;
MEDLINE=99229875; PubMed=10215036;
Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;
"Early gene expression during androgen-induced inhibition of
proliferation of prostate cancer cells: a new suppressor candidate on
chromosome 13, in the BRCA2-Rb1 locus.";
J. Steroid Biochem. Mol. Biol. 68:41-45(1999).
EMBL: U95825; AAD22134.2; -;
InterPro: IPR002687; -;
ProDom: PD004104; -; 1;
SEQUENCE 1391 AA; 158035 MW; F58ABEE5AD6D9479 CRC64;

Query Match 25.4%; Score 353; DB 4; Length 1391;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 853; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 257 DIILELYNTDLSLLSVLPQLEFKLSNDNEERLQVVKLLAKMFGAKDSELSQNKPLWQ 316
DB 257 DIILELYNTDLSLLSVLPQLEFKLSNDNEERLQVVKLLAKMFGAKDSELSQNKPLWQ 316
QY 317 CYLGRFNDIHPVIRLECYKFASHCLMHPDLAKDLTEYLKVRSHDPPEAIRHDVIVSYT 376
DB 317 CYLGRFNDIHPVIRLECYKFASHCLMHPDLAKDLTEYLKVRSHDPPEAIRHDVIVSYT 376
QY 377 AAKDILLVNDHLLNFVRETRTDKKWRVRKEAMGLAQIYKYYALQSAAGKDAAKOIAWI 436
DB 377 AAKDILLVNDHLLNFVRETRTDKKWRVRKEAMGLAQIYKYYALQSAAGKDAAKOIAWI 436
QY 437 KDKLLHIYQNSIDRLLVERIFAQYMPVPHNLETTERMKCLYLYLATDLNNAVKALNEWM 496
DB 437 KDKLLHIYQNSIDRLLVERIFAQYMPVPHNLETTERMKCLYLYLATDLNNAVKALNEWM 496
QY 497 KCONLLRHQVKDLDLILQPKTDASVKAIFSKVMVITRNLPDGPRAQDPKMKFTQVLEDD 556
DB 497 KCONLLRHQVKDLDLILQPKTDASVKAIFSKVMVITRNLPDGPRAQDPKMKFTQVLEDD 556
QY 557 EKIRKQLEVLVSPTSCQAEQCVREITKGLGNPKQPTNPFLEIMKFLERLAPVHIDTE 616
DB 557 EKIRKQLEVLVSPTSCQAEQCVREITKGLGNPKQPTNPFLEIMKFLERLAPVHIDTE 616
QY 617 SISALIKQVNSIDGTADDEGVPDQAIRAGLELLKVLSTHPSFHSATFESLLAC 676
DB 617 SISALIKQVNSIDGTADDEGVPDQAIRAGLELLKVLSTHPSFHSATFESLLAC 676
QY 677 LKMDDEKVAEALQIFKNTGSIIEEDFPHRSALLPVLHKKSGKPPROAKYAIHCIIAI 736
DB 677 LKMDDEKVAEALQIFKNTGSIIEEDFPHRSALLPVLHKKSGKPPROAKYAIHCIIAI 736
QY 737 FSSKETQFAQIPEPLHKSIDPSNLEHLITPLVTIGHIALLAPDQFAAPKMSWATFIYKD 796
DB 737 FSSKETQFAQIPEPLHKSIDPSNLEHLITPLVTIGHIALLAPDQFAAPKMSWATFIYKD 796
QY 797 LLMNDRLPGKTKTKLWVPDEEVSPETWVKIQAKMVRWLLGMKNHNSKSGTSLRLTT 856
DB 797 LLMNDRLPGKTKTKLWVPDEEVSPETWVKIQAKMVRWLLGMKNHNSKSGTSLRLTT 856
QY 857 ILHSDGDLTEQCKISKPDMSRLRGAAGSIVKLAQPCVHEITILEQYOLCALAINDECY 916
DB 857 ILHSDGDLTEQCKISKPDMSRLRGAAGSIVKLAQPCVHEITILEQYOLCALAINDECY 916
QY 917 QYRQVFAQKLHGLSLRLPLEYMAICALCAKDPVKERRAHQCLVKNINVRREYLKQH 976
DB 917 QYRQVFAQKLHGLSLRLPLEYMAICALCAKDPVKERRAHQCLVKNINVRREYLKQH 976
QY 977 AAVSEKLLSLLPEYVPTTHLLAHDPPYVKVDIEQLKDVKECLWFLVEILMAKNENNS 1036
DB 977 AAVSEKLLSLLPEYVPTTHLLAHDPPYVKVDIEQLKDVKECLWFLVEILMAKNENNS 1036
QY 1037 HAFIRKVENIKQTKDAQGPDADKMKNEKLYTVCDVAMNIMSKSTYSLESKDPVLPAR 1096

Db 1037 HAFIRKVENIKQTKDAQGPDADKMKNEKLYTVCDVAMNIMSKSTYSLESKDPVLPAR 1096
QY 1097 FFTQDPKNSNTKNYLPP 1114
Db 1097 FFTQDPKNSNTKNYLPP 1114
RESULT 4
Q9H5N8
ID Q9H5N8 PRELIMINARY; PRT; 125 AA.
AC Q9H5N8
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CDNA: FLJ23236 FIS; CLONE COL00725.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK026889; BAB15584.1; -;
SQ SEQUENCE 125 AA; 14281 MW; 492809836D3665D3 CRC64;

Query Match 9.0%; Score 125; DB 4; Length 125;
Best Local Similarity 100.0%; Pred. No. 1e-122;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1231 MDDLKLVQEQPKGSKRSKRKRGHTASESDSQWPEEKRLKEDILENEDEQNSPPKKGKR 1290
DB 1 MDDLKLVQEQPKGSKRSKRKRGHTASESDSQWPEEKRLKEDILENEDEQNSPPKKGKR 60
QY 1291 GRPPPLGGGTGKEPTMTKTSKKGSKKSGPPAPPEEEEEERQSGNTEQKSKQHRVSR 1350
DB 61 GRPPPLGGGTGKEPTMTKTSKKGSKKSGPPAPPEEEEEERQSGNTEQKSKQHRVSR 120
QY 1351 RAQR 1355
Db 121 RAQR 125

RESULT 5
Q9Y4D4
ID Q9Y4D4 PRELIMINARY; PRT; 851 AA.
AC Q9Y4D4
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE KIAA0648 PROTEIN (FRAGMENT).
GN KIAA0648.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL: AB014548; BAA31623.1; -;

1214 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGQSQRKRKRGHTASEDSDEQWPEEKRLKED 1273
 1183 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGQSQRKRKRGHTASEDSDEQWPEEKRLKED 1242
 1274 ILENEDEQNSPPKGGKRGPRPPKPLGGGTPEEPTMTSKSGKSKSGPPAPEEEEEERQ 1333
 1243 ILENEDEQNSPPKGGKRGPRPPKPLGGGTPEEPTMTSKSGKSKSGPPAPEEEEEERQ 1302
 1334 SGNTEOKSKSKQHRVSRRAQRAESPESAIESTQSTPQKGRGRPSKTPSPQPKNV 1391
 1303 SGNTEOKSKSKQHRVSRRAQRAESPESAIESTQSTPQKGRGRPSKTPSPQPKNV 1360

RESULT 2
 Q9NT15 PRELIMINARY; PRT: 1447 AA.
 AC Q9NT15;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE HYPOTHETICAL 164.7 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rhodes S., Huckle E.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AL137201; CAB69911.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1447 AA; 164666 MW; 145C30308EA3EFD5 CRC64;

Query Match - 73.0%; Score 1015; DB 4; Length 1447;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1315; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

74 KVRLLVACCLADIFRIYAPEAPYTPDKLKIDFMFTRQLKGLDTPKSPQFNRYFYLL 133
 74 KVRLLVACCLADIFRIYAPEAPYTPDKLKIDFMFTRQLKGLDTPKSPQFNRYFYLL 133
 134 NIAWKSYNICFELEDSNEIFTOLYTLFVSINNGHNVHMHVMDLMSIICEGDTVSQ 193
 134 NIAWKSYNICFELEDSNEIFTOLYTLFVSINNGHNVHMHVMDLMSIICEGDTVSQ 193
 194 ELLDTVLVNLVPAHKNLQAYDLAKALKRTAQAIPEYITTFNQLMLGKTSISDLSE 253
 194 ELLDTVLVNLVPAHKNLQAYDLAKALKRTAQAIPEYITTFNQLMLGKTSISDLSE 253
 254 HVPDLILELYNIDSHLLSVLPOLFKLSNDNEERLQVVKLLAKMFGAKDSLASQNP 313
 254 HVPDLILELYNIDSHLLSVLPOLFKLSNDNEERLQVVKLLAKMFGAKDSLASQNP 313
 314 LMOCYLGRENTHVPLRLECVKFASHCLMNHDPDLAKLTFLKVRSHDPEEARHDIIVS 373
 314 LMOCYLGRENTHVPLRLECVKFASHCLMNHDPDLAKLTFLKVRSHDPEEARHDIIVS 373
 374 IVTAARKDILLVNDHLLNFVRETLDRVRVRKEAMGLAQIYKYLQASAGKDAKQI 433
 374 IVTAARKDILLVNDHLLNFVRETLDRVRVRKEAMGLAQIYKYLQASAGKDAKQI 433
 434 AWTKDLHLIIYQNSIDRLVRIERFAQYVPHNLFTTERMKCLLYLYATLDLNAVKN 493
 434 AWTKDLHLIIYQNSIDRLVRIERFAQYVPHNLFTTERMKCLLYLYATLDLNAVKN 493
 494 EMWKNLRLHVOVKLLDLIKPKTDASVKAIFSKVAVITRNLPDPGKADFMKKTQVL 553
 494 EMWKNLRLHVOVKLLDLIKPKTDASVKAIFSKVAVITRNLPDPGKADFMKKTQVL 553
 554 EDEKIRKQLEVLVSPSCSKQAEGCVREITTKLGNPKQPTNPFLEMIKFLLERIAPVHI 613

Db 554 EDEKIRKQLEVLVSPSCSKQAEGCVREITTKLGNPKQPTNPFLEMIKFLLERIAPVHI 613
 QY 614 DTESISALIKQVKNKSIDGTADDEDEGVPTDOATRAGLELLKVLSTFTHPSFHAETFESL 673
 Db 614 DTESISALIKQVKNKSIDGTADDEDEGVPTDOATRAGLELLKVLSTFTHPSFHAETFESL 673
 QY 674 LACLKMDDEKVAEAAALQIFKNTGSKIETEDFPPIRSALLPVLHHKSKGPPROAKYAIHCI 733
 Db 674 LACLKMDDEKVAEAAALQIFKNTGSKIETEDFPPIRSALLPVLHHKSKGPPROAKYAIHCI 733
 QY 734 HAIFSKKETQFAQIFPEPLHKSIDPSNLEHLITPLVTIGHIALAPDOFAAPKWSWATFI 793
 Db 734 HAIFSKKETQFAQIFPEPLHKSIDPSNLEHLITPLVTIGHIALAPDOFAAPKWSWATFI 793
 QY 794 VKDLLMNDRLPGKKTTLKLVPPDEEVSPTWVKIQATKMMVRLGLGNHNSKSGTSTLRL 853
 Db 794 VKDLLMNDRLPGKKTTLKLVPPDEEVSPTWVKIQATKMMVRLGLGNHNSKSGTSTLRL 853
 QY 854 LTTILHSDGDLTEOGKISKPDMSRLRLAAGSAIVKLAQPCPYHEIITLEQYQCALAIND 913
 Db 854 LTTILHSDGDLTEOGKISKPDMSRLRLAAGSAIVKLAQPCPYHEIITLEQYQCALAIND 913
 QY 914 ECVQVQVFAQKLHKLGLSRLRPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYL 973
 Db 914 ECVQVQVFAQKLHKLGLSRLRPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYL 973
 QY 974 KQHAAYSEKLLSLLPEYVVPYTHLLAHPDYPYKQDIEQLKDVKECLFWLEILMAKNE 1033
 Db 974 KQHAAYSEKLLSLLPEYVVPYTHLLAHPDYPYKQDIEQLKDVKECLFWLEILMAKNE 1033
 QY 1034 NNSHAFIRKMWENIKOTKDAQGDDAKMNEKLYTCDVAMNIIKMSKSTTYSLSPKDPVL 1093
 Db 1034 NNSHAFIRKMWENIKOTKDAQGDDAKMNEKLYTCDVAMNIIKMSKSTTYSLSPKDPVL 1093
 QY 1094 PARFFTPQDNFNTKNYLPPEMKSEFTFGPKKTTNVLGAVNKPLSSACKQSQTSSRME 1153
 Db 1094 PARFFTPQDNFNTKNYLPPEMKSEFTFGPKKTTNVLGAVNKPLSSACKQSQTSSRME 1153
 QY 1154 TVSNASSSNPSPGRIKGLDSEMDHSENYDTWSSPLPGKSKDKRDSDLVRSLEK 1213
 Db 1154 TVSNASSSNPSPGRIKGLDSEMDHSENYDTWSSPLPGKSKDKRDSDLVRSLEK 1213
 QY 1214 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGQSQRKRKRGHTASEDSDEQWPEEKRLKED 1273
 Db 1214 PRGRKKTPTVEQEKLGMDLTKLVQEQPKGQSQRKRKRGHTASEDSDEQWPEEKRLKED 1273
 QY 1274 ILENEDEQNSPPKGGKRGPRPPKPLGGGTPEEPTMTSKSGKSKSGPPAPEEEEEERQ 1333
 Db 1274 ILENEDEQNSPPKGGKRGPRPPKPLGGGTPEEPTMTSKSGKSKSGPPAPEEEEEERQ 1333
 QY 1334 SGNTEOKSKSKQHRVSRRAQRAESPESAIESTQSTPQKGRGRPSKTPSPQPKNV 1391
 Db 1334 SGNTEOKSKSKQHRVSRRAQRAESPESAIESTQSTPQKGRGRPSKTPSPQPKNV 1391

RESULT 3
 Q9Y451 PRELIMINARY; PRT: 1391 AA.
 AC Q9Y451;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN.
 GN AS3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;
 RT "Androgen-induced proliferative shutoff in prostate cancer cells.";


```

969 7 0.5 719 11 Q61281 mus musculus
970 7 0.5 720 3 O14367 schizosacch
971 7 0.5 721 5 Q9NAB9 caenorhabdi
972 7 0.5 722 5 Q9VFK4 drosophila
973 7 0.5 725 4 Q9UFA6 homo sapien
974 7 0.5 725 4 Q9H986 homo sapien
975 7 0.5 726 10 Q9FGN7 arabidopsis
976 7 0.5 726 11 Q9QZV4 mus musculus
977 7 0.5 730 14 Q06634 bovine herp
978 7 0.5 732 14 Q9DUC7 Q9duc7 tt virus. o
979 7 0.5 735 13 Q9DDQ9 Q9ddo9 xenopus lae
980 7 0.5 736 6 O18805 o18805 cercopithe
981 7 0.5 736 11 Q9ESN0 Q9esn0 rattus norv
982 7 0.5 737 4 Q9UP58 Q9up58 homo sapien
983 7 0.5 738 4 Q9UNY2 Q9uny2 homo sapien
984 7 0.5 738 11 Q9WVE4 Q9wve4 mus musculus
985 7 0.5 738 11 Q9QY02 Q9qy02 rattus norv
986 7 0.5 739 10 Q9FTT9 Q9ftt9 arabidopsis
987 7 0.5 739 11 Q35613 Q35613 mus musculus
988 7 0.5 740 4 O15141 O15141 homo sapien
989 7 0.5 740 4 Q9UPP2 Q9upp2 homo sapien
990 7 0.5 740 4 Q9UER7 Q9uer7 homo sapien
991 7 0.5 740 4 O14747 O14747 homo sapien
992 7 0.5 740 11 Q9QWT8 Q9qwt8 mus musculus
993 7 0.5 741 3 Q9Y7K6 Q9y7k6 schizosacch
994 7 0.5 742 5 Q9NEB3 Q9neb3 leishmania
995 7 0.5 742 10 Q9FW12 Q9fw12 oryza sativ
996 7 0.5 748 4 Q9Y6E1 Q9y6e1 homo sapien
997 7 0.5 748 4 Q9UBJ0 Q9ubj0 homo sapien
998 7 0.5 750 14 Q9Y039 Q9yq39 pseudorabie
999 7 0.5 751 5 Q917S7 Q917s7 drosophila
1000 7 0.5 751 11 Q60709 Q60709 mus musculus

```

ALIGNMENTS

RESULT 1

```

Q9Y2I5 ID Q9Y2I5 PRELIMINARY; PRT; 1416 AA.
AC Q9Y2I5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE KIAA0979 PROTEIN (FRAGMENT).
GN KIAA0979.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
SEQUENCE FROM N.A.
RE TISSUE=BRAIN;
RC MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023196; BAA76823.1; -.
DR InterPro; IPR000637; -.
DR PRINTS; PR00929; ATHOOK.
FT NON_TER 1
SQ SEQUENCE 1416 AA; 161192 MW; 0BC682D7F4657424 CRC64;

```

```

Query Match 73.0%; Score 1015; DB 4; Length 1416;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1315; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

74 KDVRLLVACCLADIPRIYAPEAPYTPDKLIDIFWFIITROLKGLDTSKSPQNRVYFILLE 133
|||||

```

```

Db 43 KDVRLLVACCLADIPRIYAPEAPYTPDKLIDIFWFIITROLKGLDTSKSPQNRVYFILLE 102
QY 134 NIAWKSYNICFELEDSNEIETQLYRTLFVSINNGHNOKVHMHVMDLMSIICEGDTVSQ 193
Db 103 NIAWKSYNICFELEDSNEIETQLYRTLFVSINNGHNOKVHMHVMDLMSIICEGDTVSQ 162
QY 194 ELLDVLVNLPAHKNLKNQAYDLAKALLKRTAQAIETPIITFFNOVLMGLKTSISDLS 253
Db 163 ELLDVLVNLPAHKNLKNQAYDLAKALLKRTAQAIETPIITFFNOVLMGLKTSISDLS 222
QY 254 HVFDLILELYNIDSHLLSVLPOLFEKLSNDNERLQVVKLLAKMFGAKDSELAONKP 313
Db 223 HVFDLILELYNIDSHLLSVLPOLFEKLSNDNERLQVVKLLAKMFGAKDSELAONKP 282
QY 314 LWQCYLGRFNIDHVPILRELCVKFASHCLMNHDPDLAKDLTEYLKVRSHDPEEAIHRDVI 373
Db 283 LWQCYLGRFNIDHVPILRELCVKFASHCLMNHDPDLAKDLTEYLKVRSHDPEEAIHRDVI 342
QY 374 IVTAAKDDILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYALQSAAGKDAAKQI 433
Db 343 IVTAAKDDILLVNDHLLNFVRERTLDKRWVRKEAMGLAQIYKKYALQSAAGKDAAKQI 402
QY 434 AWIKDKLLHIYQNSIDRDLVERIFAQVMVPHNLETTERMKCLYLYATLDLNAVKA 493
Db 403 AWIKDKLLHIYQNSIDRDLVERIFAQVMVPHNLETTERMKCLYLYATLDLNAVKA 462
QY 494 EMWQCNLLRHQVKDLLDIKOPKTASVKAIFSVMVITRNLPDGPKAQDPKMKFTQVL 553
Db 463 EMWQCNLLRHQVKDLLDIKOPKTASVKAIFSVMVITRNLPDGPKAQDPKMKFTQVL 522
QY 554 EDEKIRKQLEVLVSPTCSQAQECVREITKKNLPQPTNPFLKILERTAPVHI 613
Db 523 EDEKIRKQLEVLVSPTCSQAQECVREITKKNLPQPTNPFLKILERTAPVHI 582
QY 614 DTESIALIKOVNSIDGTADDEGVPTDQAIRAGLELLKVLSTHPTSFHSAETFESL 673
Db 583 DTESIALIKOVNSIDGTADDEGVPTDQAIRAGLELLKVLSTHPTSFHSAETFESL 642
QY 674 LACLKMDDEKVAEAAALQIFKNTGSKIEEDFPHIRSALLPVLHKKSGKPPQAKYAIHCI 733
Db 643 LACLKMDDEKVAEAAALQIFKNTGSKIEEDFPHIRSALLPVLHKKSGKPPQAKYAIHCI 702
QY 734 HAIFSSKETQFAQIEPEPLHKSLDPSNLEHLITPLVTIGHIALLAPDOFAAPKSWVATPI 793
Db 703 HAIFSSKETQFAQIEPEPLHKSLDPSNLEHLITPLVTIGHIALLAPDOFAAPKSWVATPI 762
QY 794 VKDLMLNDRLPGKTKTKLWVPDEEVSPTMWKIQAIKMMVRLLGMKNHNSKSGTSTLRL 853
Db 763 VKDLMLNDRLPGKTKTKLWVPDEEVSPTMWKIQAIKMMVRLLGMKNHNSKSGTSTLRL 822
QY 854 LTTILHSDGDLTEQCKISKPDMSRLRLAAGSAIVKLAQEPYHEIITLQYQCALAIND 913
Db 823 LTTILHSDGDLTEQCKISKPDMSRLRLAAGSAIVKLAQEPYHEIITLQYQCALAIND 882
QY 914 ECVQYRVQFAQKLHKLRLPLEYMAICALKADPKVKEHRAHARQCLVKNNVREYL 973
Db 883 ECVQYRVQFAQKLHKLRLPLEYMAICALKADPKVKEHRAHARQCLVKNNVREYL 942
QY 974 KOHAAVSEKLLSLPEYVVPYTIHLLAHDPDYVKVDTEQLDKVKECLWFLVLEILMAKNE 1033
Db 943 KOHAAVSEKLLSLPEYVVPYTIHLLAHDPDYVKVDTEQLDKVKECLWFLVLEILMAKNE 1002
QY 1034 NNSHAFIRKMWENIKQTKDAQOPDDAKMNEKLYTVCDVAMNIIIMSKSTYTSLESKDPVL 1093
Db 1003 NNSHAFIRKMWENIKQTKDAQOPDDAKMNEKLYTVCDVAMNIIIMSKSTYTSLESKDPVL 1062
QY 1094 PARFTQDPKNFNTKNYLPPEMKSFPTPGPKTTNVLGAVNKPPLSSAGKQSQTSRME 1153
Db 1063 PARFTQDPKNFNTKNYLPPEMKSFPTPGPKTTNVLGAVNKPPLSSAGKQSQTSRME 1122
QY 1154 TVSNASSSSNPPSGRIKGRLDSSMDHSENEDEYTMSSPLPGCKSKDKRDDSDLVVRSELEK 1213
Db 1123 TVSNASSSSNPPSGRIKGRLDSSMDHSENEDEYTMSSPLPGCKSKDKRDDSDLVVRSELEK 1182

```


| | | | | | | | | | | | | | |
|-----|---|-----|-----|----|--------|---------------------|-----|---|-----|-----|----|--------|---------------------|
| 823 | 7 | 0.5 | 580 | 10 | Q9ZWB0 | Q9Zwb0 arabidopsis | 896 | 7 | 0.5 | 634 | 5 | Q06279 | Q06279 schistosoma |
| 824 | 7 | 0.5 | 580 | 10 | Q9M2P6 | Q9m2p6 arabidopsis | 897 | 7 | 0.5 | 635 | 2 | P71012 | P71012 bacillus su |
| 825 | 7 | 0.5 | 580 | 11 | Q98291 | Q98291 mus musculus | 898 | 7 | 0.5 | 635 | 5 | Q13476 | Q13476 homo sapien |
| 826 | 7 | 0.5 | 580 | 11 | Q9EQR3 | Q9eqr3 mus musculus | 899 | 7 | 0.5 | 637 | 5 | P91656 | P91656 drosophila |
| 827 | 7 | 0.5 | 581 | 5 | Q9U4P1 | Q9u4p1 arabidopsis | 900 | 7 | 0.5 | 639 | 3 | Q43113 | Q43113 neurospora |
| 828 | 7 | 0.5 | 581 | 10 | Q9SX39 | Q9sx39 arabidopsis | 901 | 7 | 0.5 | 639 | 4 | Q9H8K3 | Q9h8k3 homo sapien |
| 829 | 7 | 0.5 | 582 | 5 | Q23909 | Q23909 dictyostell | 902 | 7 | 0.5 | 639 | 10 | Q9LN44 | Q9ln44 arabidopsis |
| 830 | 7 | 0.5 | 582 | 5 | Q9U4P9 | Q9u4p9 drosophila | 903 | 7 | 0.5 | 642 | 2 | Q9FC03 | Q9fc03 streptomyce |
| 831 | 7 | 0.5 | 582 | 5 | Q9U4P8 | Q9u4p8 drosophila | 904 | 7 | 0.5 | 642 | 5 | Q93894 | Q93894 caenorhabdi |
| 832 | 7 | 0.5 | 582 | 5 | Q9U4P7 | Q9u4p7 drosophila | 905 | 7 | 0.5 | 646 | 10 | Q9XEC7 | Q9xec7 arabidopsis |
| 833 | 7 | 0.5 | 582 | 5 | Q9U4P3 | Q9u4p3 drosophila | 906 | 7 | 0.5 | 646 | 11 | Q63802 | Q63802 rattus norv |
| 834 | 7 | 0.5 | 584 | 5 | Q21088 | Q21088 caenorhabdi | 907 | 7 | 0.5 | 646 | 11 | Q9EPL7 | Q9epL7 mus musculus |
| 835 | 7 | 0.5 | 584 | 5 | Q9U4P6 | Q9u4p6 drosophila | 908 | 7 | 0.5 | 647 | 14 | Q9DWB9 | Q9dwb9 rat cytomeg |
| 836 | 7 | 0.5 | 584 | 5 | Q9U4P5 | Q9u4p5 drosophila | 909 | 7 | 0.5 | 648 | 2 | Q31644 | Q31644 bacillus su |
| 837 | 7 | 0.5 | 584 | 5 | Q9U4P2 | Q9u4p2 drosophila | 910 | 7 | 0.5 | 649 | 3 | Q60167 | Q60167 schizosacch |
| 838 | 7 | 0.5 | 584 | 11 | Q9EQ61 | Q9eq61 mus musculus | 911 | 7 | 0.5 | 649 | 4 | Q9E215 | Q9e215 homo sapien |
| 839 | 7 | 0.5 | 585 | 3 | Q9HDV2 | Q9hdv2 schizosacch | 912 | 7 | 0.5 | 650 | 11 | P70663 | P70663 mus musculus |
| 840 | 7 | 0.5 | 585 | 10 | Q9LMD2 | Q9lmd2 arabidopsis | 913 | 7 | 0.5 | 651 | 5 | Q18181 | Q18181 caenorhabdi |
| 841 | 7 | 0.5 | 587 | 2 | Q9HUC8 | Q9huc8 pseudomonas | 914 | 7 | 0.5 | 651 | 14 | Q9WNM8 | Q9wnm8 human papil |
| 842 | 7 | 0.5 | 588 | 5 | Q16044 | Q16044 drosophila | 915 | 7 | 0.5 | 652 | 4 | Q9Y6N9 | Q9y6n9 homo sapien |
| 843 | 7 | 0.5 | 588 | 5 | Q9VHJ3 | Q9vhj3 drosophila | 916 | 7 | 0.5 | 654 | 4 | Q9NW82 | Q9nw82 homo sapien |
| 844 | 7 | 0.5 | 590 | 2 | Q9KXF5 | Q9kyf5 streptomyce | 917 | 7 | 0.5 | 656 | 11 | Q61148 | Q61148 mus musculus |
| 845 | 7 | 0.5 | 590 | 4 | Q9H2P2 | Q9h2p2 homo sapien | 918 | 7 | 0.5 | 661 | 2 | Q9JZF3 | Q9jzf3 neisseria m |
| 846 | 7 | 0.5 | 592 | 5 | Q9VE48 | Q9ve48 drosophila | 919 | 7 | 0.5 | 661 | 5 | Q22249 | Q22249 caenorhabdi |
| 847 | 7 | 0.5 | 593 | 3 | Q9WU74 | Q9wu74 rattus norv | 920 | 7 | 0.5 | 662 | 3 | Q9HFB3 | Q9hfb3 aspergillus |
| 848 | 7 | 0.5 | 594 | 11 | Q9EQI9 | Q9eqi9 mus musculus | 921 | 7 | 0.5 | 663 | 10 | Q9XEP8 | Q9xep8 sorghum bic |
| 849 | 7 | 0.5 | 594 | 13 | Q9DF69 | Q9df69 gallus gall | 922 | 7 | 0.5 | 664 | 2 | Q05871 | Q05871 mycobacteri |
| 850 | 7 | 0.5 | 595 | 4 | Q9UEU4 | Q9ueu4 homo sapien | 923 | 7 | 0.5 | 664 | 13 | Q9IAT6 | Q9iat6 brachydanio |
| 851 | 7 | 0.5 | 596 | 5 | Q45633 | Q45633 caenorhabdi | 924 | 7 | 0.5 | 667 | 5 | Q9NM60 | Q9nm60 caenorhabdi |
| 852 | 7 | 0.5 | 597 | 4 | Q12839 | Q12839 homo sapien | 925 | 7 | 0.5 | 667 | 11 | Q35745 | Q35745 mus musculus |
| 853 | 7 | 0.5 | 598 | 4 | Q13387 | Q13387 homo sapien | 926 | 7 | 0.5 | 668 | 10 | Q9SVM5 | Q9svm5 arabidopsis |
| 854 | 7 | 0.5 | 599 | 5 | Q97009 | Q97009 leishmania | 927 | 7 | 0.5 | 669 | 4 | Q14662 | Q14662 homo sapien |
| 855 | 7 | 0.5 | 600 | 4 | Q9NSB2 | Q9nsb2 homo sapien | 928 | 7 | 0.5 | 671 | 4 | Q9HAK1 | Q9hak1 homo sapien |
| 856 | 7 | 0.5 | 601 | 3 | Q9HE54 | Q9he54 neurospora | 929 | 7 | 0.5 | 671 | 4 | Q9H4G3 | Q9h4g3 homo sapien |
| 857 | 7 | 0.5 | 602 | 4 | Q9HA98 | Q9ha98 homo sapien | 930 | 7 | 0.5 | 672 | 5 | Q44189 | Q44189 caenorhabdi |
| 858 | 7 | 0.5 | 603 | 2 | Q9EYV2 | Q9eyv2 rhizobium l | 931 | 7 | 0.5 | 673 | 5 | Q9U0N1 | Q9u0n1 plasmodium |
| 859 | 7 | 0.5 | 603 | 8 | Q9Y907 | Q9y907 alligator m | 932 | 7 | 0.5 | 673 | 11 | Q9R0Q1 | Q9r0q1 mus musculus |
| 860 | 7 | 0.5 | 603 | 8 | Q33790 | Q33790 alligator m | 933 | 7 | 0.5 | 674 | 2 | Q9KQ91 | Q9kq91 vibrio chol |
| 861 | 7 | 0.5 | 603 | 10 | Q22787 | Q22787 arabidopsis | 934 | 7 | 0.5 | 674 | 4 | Q9H053 | Q9h053 homo sapien |
| 862 | 7 | 0.5 | 603 | 10 | Q9ST93 | Q9st93 oryza sativ | 935 | 7 | 0.5 | 675 | 5 | Q61310 | Q61310 trichineila |
| 863 | 7 | 0.5 | 605 | 3 | Q9URJ8 | Q9urj8 penicillium | 936 | 7 | 0.5 | 675 | 11 | Q9EPT3 | Q9ept3 mus musculus |
| 864 | 7 | 0.5 | 605 | 3 | Q9NSK3 | Q9nsk3 caenorhabdi | 937 | 7 | 0.5 | 676 | 10 | Q22762 | Q22762 arabidopsis |
| 865 | 7 | 0.5 | 607 | 5 | Q9VHA9 | Q9vha9 drosophila | 938 | 7 | 0.5 | 678 | 3 | Q94235 | Q94235 schizosacch |
| 866 | 7 | 0.5 | 608 | 5 | Q9V6D4 | Q9v6d4 drosophila | 939 | 7 | 0.5 | 679 | 2 | Q916V6 | Q9i6v6 pseudomonas |
| 867 | 7 | 0.5 | 608 | 10 | Q9SUL2 | Q9sul2 arabidopsis | 940 | 7 | 0.5 | 680 | 5 | Q18496 | Q18496 caenorhabdi |
| 868 | 7 | 0.5 | 608 | 11 | Q9QVN9 | Q9qvn9 cricetus | 941 | 7 | 0.5 | 681 | 5 | Q9V8K2 | Q9v8k2 drosophila |
| 869 | 7 | 0.5 | 610 | 3 | Q12136 | Q12136 saccharomyc | 942 | 7 | 0.5 | 682 | 5 | Q9GTW3 | Q9gtw3 plasmodium |
| 870 | 7 | 0.5 | 610 | 10 | Q9FXE6 | Q9fxe6 arabidopsis | 943 | 7 | 0.5 | 683 | 10 | Q23506 | Q23506 arabidopsis |
| 871 | 7 | 0.5 | 612 | 5 | Q9TX06 | Q9tx06 dictyostell | 944 | 7 | 0.5 | 684 | 5 | Q9VXS2 | Q9vxs2 drosophila |
| 872 | 7 | 0.5 | 614 | 5 | P91956 | P91956 lytechinus | 945 | 7 | 0.5 | 685 | 10 | Q9SS80 | Q9ss80 arabidopsis |
| 873 | 7 | 0.5 | 618 | 10 | Q9M4B7 | Q9m4b7 arabidopsis | 946 | 7 | 0.5 | 687 | 11 | Q89024 | Q89024 mus musculus |
| 874 | 7 | 0.5 | 618 | 10 | Q9LKL2 | Q9lkl2 arabidopsis | 947 | 7 | 0.5 | 687 | 11 | Q9WUB6 | Q9wub6 mus musculus |
| 875 | 7 | 0.5 | 618 | 14 | Q9QM78 | Q9qm78 avian adeno | 948 | 7 | 0.5 | 688 | 11 | Q88934 | Q88934 mus musculus |
| 876 | 7 | 0.5 | 619 | 2 | Q9X4G3 | Q9x4g3 aquifex pyr | 949 | 7 | 0.5 | 691 | 13 | Q57658 | Q57658 gallus gall |
| 877 | 7 | 0.5 | 619 | 4 | Q95358 | Q95358 homo sapien | 950 | 7 | 0.5 | 693 | 14 | Q9VQ38 | Q9vq38 pseudorabie |
| 878 | 7 | 0.5 | 619 | 4 | Q9HDC6 | Q9hdc6 homo sapien | 951 | 7 | 0.5 | 694 | 4 | Q9X2J0 | Q9x2j0 homo sapien |
| 879 | 7 | 0.5 | 619 | 10 | Q9SH81 | Q9sh81 arabidopsis | 952 | 7 | 0.5 | 694 | 10 | Q9XFS2 | Q9xfs2 arabidopsis |
| 880 | 7 | 0.5 | 619 | 10 | Q9LV37 | Q9lv37 arabidopsis | 953 | 7 | 0.5 | 695 | 4 | Q13861 | Q13861 homo sapien |
| 881 | 7 | 0.5 | 619 | 14 | Q73453 | Q73453 human papil | 954 | 7 | 0.5 | 695 | 11 | Q64348 | Q64348 mus musculus |
| 882 | 7 | 0.5 | 621 | 2 | Q9ZFR4 | Q9zfr4 burkholderi | 955 | 7 | 0.5 | 696 | 4 | Q9UJK6 | Q9ujk6 homo sapien |
| 883 | 7 | 0.5 | 622 | 13 | Q98985 | Q98985 rana rugosa | 956 | 7 | 0.5 | 696 | 4 | Q00461 | Q00461 homo sapien |
| 884 | 7 | 0.5 | 623 | 5 | Q9W4A2 | Q9w4a2 drosophila | 957 | 7 | 0.5 | 700 | 4 | Q75054 | Q75054 homo sapien |
| 885 | 7 | 0.5 | 625 | 2 | Q67704 | Q67704 aquifex aeo | 958 | 7 | 0.5 | 700 | 10 | Q9SVT1 | Q9svt1 arabidopsis |
| 886 | 7 | 0.5 | 626 | 10 | Q9LDL2 | Q9ldl2 arabidopsis | 959 | 7 | 0.5 | 704 | 5 | Q9VSR3 | Q9vsr3 drosophila |
| 887 | 7 | 0.5 | 627 | 10 | Q9MAS7 | Q9mas7 arabidopsis | 960 | 7 | 0.5 | 704 | 10 | Q9FF49 | Q9ff49 arabidopsis |
| 888 | 7 | 0.5 | 628 | 2 | Q84023 | Q84023 chlamydia t | 961 | 7 | 0.5 | 704 | 11 | Q9JLQ4 | Q9jlq4 mus musculus |
| 889 | 7 | 0.5 | 629 | 3 | Q9P6U6 | Q9p6u6 neurospora | 962 | 7 | 0.5 | 708 | 3 | Q06411 | Q06411 saccharomyc |
| 890 | 7 | 0.5 | 629 | 13 | Q93310 | Q93310 xenopus lae | 963 | 7 | 0.5 | 708 | 5 | Q9WLD3 | Q9wld3 drosophila |
| 891 | 7 | 0.5 | 630 | 10 | Q9LZE2 | Q9lze2 arabidopsis | 964 | 7 | 0.5 | 709 | 5 | Q9VU89 | Q9vu89 drosophila |
| 892 | 7 | 0.5 | 632 | 4 | Q9NTE0 | Q9nte0 homo sapien | 965 | 7 | 0.5 | 712 | 11 | Q54729 | Q54729 rattus norv |
| 893 | 7 | 0.5 | 632 | 4 | Q9H5L3 | Q9h5l3 homo sapien | 966 | 7 | 0.5 | 715 | 10 | Q9LD59 | Q9ld59 arabidopsis |
| 894 | 7 | 0.5 | 633 | 5 | Q61104 | Q61104 trypanosoma | 967 | 7 | 0.5 | 715 | 11 | Q9Q2X1 | Q9q2x1 rattus norv |
| 895 | 7 | 0.5 | 634 | 4 | Q15209 | Q15209 homo sapien | 968 | 7 | 0.5 | 719 | 10 | Q40916 | Q40916 pseudotsuga |

| | | | | | | | | | | | | | |
|-----|---|-----|-----|----|--------|---------------------|-----|---|-----|-----|----|--------|----------------------|
| 677 | 7 | 0.5 | 477 | 6 | 097600 | 097600 oryctolagus | 750 | 7 | 0.5 | 527 | 10 | 023389 | 023389 arabidopsis |
| 678 | 7 | 0.5 | 477 | 10 | 09SUL5 | 09sul5 arabidopsis | 751 | 7 | 0.5 | 528 | 4 | 095228 | 095228 homo sapien |
| 679 | 7 | 0.5 | 479 | 1 | 09UWR2 | 09uwr2 pyrococcus | 752 | 7 | 0.5 | 528 | 4 | 09p0w9 | 09p0w9 homo sapien |
| 680 | 7 | 0.5 | 480 | 10 | 09LSK7 | 09lsk7 arabidopsis | 753 | 7 | 0.5 | 529 | 10 | 09FMV0 | 09fmv0 arabidopsis |
| 681 | 7 | 0.5 | 483 | 11 | 063035 | 063035 rattus norv | 754 | 7 | 0.5 | 531 | 2 | 09HZ10 | 09hzi0 pseudomonas |
| 682 | 7 | 0.5 | 484 | 3 | 098287 | 098287 saccharomyc | 755 | 7 | 0.5 | 531 | 5 | 09TX24 | 09txz4 caenorhabdi |
| 683 | 7 | 0.5 | 484 | 4 | 09HC22 | 09hcg2 homo sapien | 756 | 7 | 0.5 | 531 | 10 | 09LX13 | 09lxy3 arabidopsis |
| 684 | 7 | 0.5 | 484 | 10 | 09SVJ0 | 09svj0 arabidopsis | 757 | 7 | 0.5 | 531 | 10 | 09FHT0 | 09fht0 arabidopsis |
| 685 | 7 | 0.5 | 485 | 13 | 09YGS0 | 09ygs0 gallus gall | 758 | 7 | 0.5 | 531 | 10 | 09FFQ8 | 09fff8 arabidopsis |
| 686 | 7 | 0.5 | 486 | 3 | 007376 | 007376 saccharomyc | 759 | 7 | 0.5 | 532 | 5 | 09VC95 | 09vc95 drosophila |
| 687 | 7 | 0.5 | 486 | 5 | 09NKW8 | 09nk8 patinopecte | 760 | 7 | 0.5 | 533 | 3 | 001617 | 001617 caenorhabdi |
| 688 | 7 | 0.5 | 486 | 10 | 004591 | 004591 arabidopsis | 761 | 7 | 0.5 | 533 | 3 | 006631 | 006631 saccharomyc |
| 689 | 7 | 0.5 | 486 | 10 | 09FRW0 | 09frw0 arabidopsis | 762 | 7 | 0.5 | 536 | 11 | 070606 | 070606 rattus norv |
| 690 | 7 | 0.5 | 487 | 4 | 09NQ10 | 09nql0 homo sapien | 763 | 7 | 0.5 | 538 | 4 | 09NW07 | 09nwq7 homo sapien |
| 691 | 7 | 0.5 | 487 | 11 | 09JLK9 | 09jlk9 mus musculus | 764 | 7 | 0.5 | 538 | 11 | 09ET15 | 09et15 mus musculus |
| 692 | 7 | 0.5 | 490 | 4 | 09HBE6 | 09hbe6 homo sapien | 765 | 7 | 0.5 | 539 | 4 | 012899 | 012899 homo sapien |
| 693 | 7 | 0.5 | 491 | 4 | 09H4R1 | 09h4r1 homo sapien | 766 | 7 | 0.5 | 539 | 4 | 095196 | 095196 homo sapien |
| 694 | 7 | 0.5 | 491 | 10 | 09SKI6 | 09ski6 arabidopsis | 767 | 7 | 0.5 | 539 | 11 | 09QX32 | 09qxy32 mus musculus |
| 695 | 7 | 0.5 | 492 | 6 | 046384 | 046384 bos taurus | 768 | 7 | 0.5 | 540 | 4 | 075864 | 075864 homo sapien |
| 696 | 7 | 0.5 | 492 | 11 | 09Z1K6 | 09z1k6 mus musculus | 769 | 7 | 0.5 | 540 | 11 | 062253 | 062253 mus musculus |
| 697 | 7 | 0.5 | 493 | 2 | 09JMM8 | 09jmm8 wolbachia s | 770 | 7 | 0.5 | 541 | 3 | 09PE74 | 09pet4 neuospora |
| 698 | 7 | 0.5 | 493 | 4 | 095376 | 095376 homo sapien | 771 | 7 | 0.5 | 544 | 4 | 09P0J8 | 09p0j8 homo sapien |
| 699 | 7 | 0.5 | 493 | 4 | 09UEW9 | 09uem9 homo sapien | 772 | 7 | 0.5 | 544 | 11 | 062831 | 062831 rattus norv |
| 700 | 7 | 0.5 | 493 | 10 | 093015 | 093015 arabidopsis | 773 | 7 | 0.5 | 545 | 3 | 005842 | 005842 saccharomyc |
| 701 | 7 | 0.5 | 493 | 10 | 09SKS9 | 09sk9 arabidopsis | 774 | 7 | 0.5 | 545 | 6 | 077666 | 077666 sus scrofa |
| 702 | 7 | 0.5 | 494 | 3 | 09HEL2 | 09hel2 neuospora | 775 | 7 | 0.5 | 545 | 10 | 09M9V8 | 09m9v8 arabidopsis |
| 703 | 7 | 0.5 | 495 | 10 | 023330 | 023330 arabidopsis | 776 | 7 | 0.5 | 545 | 11 | 088836 | 088836 mus musculus |
| 704 | 7 | 0.5 | 496 | 4 | 09Y2K7 | 09y2k7 homo sapien | 777 | 7 | 0.5 | 547 | 10 | 09SD74 | 09sd74 arabidopsis |
| 705 | 7 | 0.5 | 497 | 11 | 09J110 | 09j110 mus musculus | 778 | 7 | 0.5 | 547 | 11 | 09ES65 | 09es65 mus musculus |
| 706 | 7 | 0.5 | 498 | 10 | 093483 | 093483 calocedrus | 779 | 7 | 0.5 | 549 | 5 | 09N5M7 | 09n5m7 caenorhabdi |
| 707 | 7 | 0.5 | 498 | 10 | 09L2C8 | 09lzc8 arabidopsis | 780 | 7 | 0.5 | 549 | 13 | 09DGM7 | 09dgm7 ictalurus p |
| 708 | 7 | 0.5 | 499 | 5 | 09Y120 | 09y120 drosophila | 781 | 7 | 0.5 | 550 | 5 | 09VD69 | 09vd69 drosophila |
| 709 | 7 | 0.5 | 499 | 10 | 09FYW3 | 09fyw3 lycopersico | 782 | 7 | 0.5 | 550 | 10 | 082347 | 082347 arabidopsis |
| 710 | 7 | 0.5 | 500 | 4 | 09UN16 | 09un16 homo sapien | 783 | 7 | 0.5 | 551 | 4 | 013512 | 013512 homo sapien |
| 711 | 7 | 0.5 | 500 | 5 | 022469 | 022469 caenorhabdi | 784 | 7 | 0.5 | 551 | 4 | 015740 | 015740 homo sapien |
| 712 | 7 | 0.5 | 501 | 2 | 09ZJ03 | 09zj03 streptococc | 785 | 7 | 0.5 | 551 | 13 | 013003 | 013003 gallus gall |
| 713 | 7 | 0.5 | 502 | 10 | 09LXH2 | 09lyh2 arabidopsis | 786 | 7 | 0.5 | 552 | 2 | 067593 | 067593 aquifex aeo |
| 714 | 7 | 0.5 | 502 | 10 | 065237 | 065237 arabidopsis | 787 | 7 | 0.5 | 552 | 2 | 09K5M6 | 09k5m6 anabaena sp |
| 715 | 7 | 0.5 | 502 | 11 | 063892 | 063892 rattus norv | 788 | 7 | 0.5 | 552 | 4 | 09UM04 | 09um04 homo sapien |
| 716 | 7 | 0.5 | 502 | 11 | 09RQ00 | 09rq00 mus musculus | 789 | 7 | 0.5 | 552 | 5 | 046178 | 046178 strongyloce |
| 717 | 7 | 0.5 | 503 | 5 | 09VL48 | 09vl48 drosophila | 790 | 7 | 0.5 | 552 | 5 | 045117 | 045117 chironomus |
| 718 | 7 | 0.5 | 504 | 5 | 09VFM0 | 09vfm0 drosophila | 791 | 7 | 0.5 | 552 | 10 | 09LIS7 | 09lis7 oryza sativ |
| 719 | 7 | 0.5 | 504 | 10 | 09SUE9 | 09sue9 arabidopsis | 792 | 7 | 0.5 | 553 | 10 | 049931 | 049931 pisum sativ |
| 720 | 7 | 0.5 | 505 | 11 | 060569 | 060569 mus cookii. | 793 | 7 | 0.5 | 553 | 11 | 09JHT9 | 09jht9 mus musculus |
| 721 | 7 | 0.5 | 507 | 11 | 054716 | 054716 mus musculus | 794 | 7 | 0.5 | 558 | 5 | 026631 | 026631 strongyloce |
| 722 | 7 | 0.5 | 507 | 11 | 035728 | 035728 mus musculus | 795 | 7 | 0.5 | 558 | 10 | 09M8S2 | 09m8s2 arabidopsis |
| 723 | 7 | 0.5 | 508 | 4 | 015208 | 015208 homo sapien | 796 | 7 | 0.5 | 559 | 10 | 09M016 | 09m016 arabidopsis |
| 724 | 7 | 0.5 | 508 | 10 | 093484 | 093484 calocedrus | 797 | 7 | 0.5 | 560 | 4 | 09NY90 | 09ny90 homo sapien |
| 725 | 7 | 0.5 | 509 | 10 | 040870 | 040870 picea glauc | 798 | 7 | 0.5 | 560 | 6 | 046411 | 046411 bos taurus |
| 726 | 7 | 0.5 | 510 | 4 | 075926 | 075926 homo sapien | 799 | 7 | 0.5 | 560 | 10 | 064883 | 064883 arabidopsis |
| 727 | 7 | 0.5 | 514 | 4 | 015223 | 015223 homo sapien | 800 | 7 | 0.5 | 562 | 4 | 012820 | 012820 homo sapien |
| 728 | 7 | 0.5 | 515 | 5 | 09V8M3 | 09v8m3 drosophila | 801 | 7 | 0.5 | 564 | 5 | 09VRX8 | 09vrx8 drosophila |
| 729 | 7 | 0.5 | 515 | 6 | 09GL76 | 09gl76 sus scrofa | 802 | 7 | 0.5 | 564 | 5 | 09V719 | 09v719 drosophila |
| 730 | 7 | 0.5 | 515 | 11 | 09JXF6 | 09jxf6 mus musculus | 803 | 7 | 0.5 | 565 | 4 | 09UB01 | 09ubq1 homo sapien |
| 731 | 7 | 0.5 | 515 | 11 | 09ERL5 | 09erl5 mus musculus | 804 | 7 | 0.5 | 565 | 4 | 09NZS5 | 09nzs5 homo sapien |
| 732 | 7 | 0.5 | 516 | 10 | 09LZS1 | 09lzs1 arabidopsis | 805 | 7 | 0.5 | 565 | 5 | 090625 | 090625 cryptospori |
| 733 | 7 | 0.5 | 516 | 11 | 09J117 | 09j117 mus musculus | 806 | 7 | 0.5 | 565 | 6 | 09TU19 | 09t19 canis famli |
| 734 | 7 | 0.5 | 517 | 4 | 075465 | 075465 homo sapien | 807 | 7 | 0.5 | 566 | 5 | 09W243 | 09w243 drosophila |
| 735 | 7 | 0.5 | 517 | 5 | 017941 | 017941 caenorhabdi | 808 | 7 | 0.5 | 566 | 10 | 024294 | 024294 pisum sativ |
| 736 | 7 | 0.5 | 517 | 5 | 09V921 | 09v921 drosophila | 809 | 7 | 0.5 | 567 | 4 | 09HD28 | 09hd28 homo sapien |
| 737 | 7 | 0.5 | 519 | 4 | 09Y508 | 09y508 homo sapien | 810 | 7 | 0.5 | 567 | 4 | 09H606 | 09h6q6 homo sapien |
| 738 | 7 | 0.5 | 519 | 4 | 09H4P2 | 09h4p2 homo sapien | 811 | 7 | 0.5 | 568 | 11 | 09EP97 | 09ep97 mus musculus |
| 739 | 7 | 0.5 | 521 | 4 | 09UM17 | 09um17 homo sapien | 812 | 7 | 0.5 | 569 | 10 | 043468 | 043468 glycine max |
| 740 | 7 | 0.5 | 522 | 10 | 09L1W0 | 09liw0 oryza sativ | 813 | 7 | 0.5 | 571 | 11 | 09ERQ6 | 09erq6 rattus norv |
| 741 | 7 | 0.5 | 523 | 3 | 008831 | 008831 saccharomyc | 814 | 7 | 0.5 | 572 | 5 | 020715 | 020715 caenorhabdi |
| 742 | 7 | 0.5 | 523 | 10 | 09LJZ2 | 09ljz2 arabidopsis | 815 | 7 | 0.5 | 573 | 4 | 09P2A5 | 09p2a5 homo sapien |
| 743 | 7 | 0.5 | 523 | 13 | 093482 | 093482 gallus gall | 816 | 7 | 0.5 | 574 | 4 | 09NXT6 | 09nxt6 homo sapien |
| 744 | 7 | 0.5 | 524 | 10 | 09LME6 | 09lme6 arabidopsis | 817 | 7 | 0.5 | 574 | 4 | 09H4L4 | 09h4l4 homo sapien |
| 745 | 7 | 0.5 | 524 | 10 | 09FJY5 | 09fjy5 arabidopsis | 818 | 7 | 0.5 | 574 | 11 | 09WU75 | 09wu75 rattus norv |
| 746 | 7 | 0.5 | 525 | 11 | 09WU76 | 09wu76 rattus norv | 819 | 7 | 0.5 | 576 | 10 | 09LUY5 | 09luuy5 arabidopsis |
| 747 | 7 | 0.5 | 526 | 2 | 09RQ57 | 09rqs7 mycoplasma | 820 | 7 | 0.5 | 578 | 5 | 09U4P4 | 09u4p4 drosophila |
| 748 | 7 | 0.5 | 526 | 2 | 091365 | 091365 pseudomonas | 821 | 7 | 0.5 | 580 | 4 | 09UF71 | 09uf71 homo sapien |
| 749 | 7 | 0.5 | 526 | 4 | 09UP59 | 09up59 homo sapien | 822 | 7 | 0.5 | 580 | 5 | 09TW69 | 09tw69 drosophila |

| | | | | | | | | | | | | | |
|-----|---|-----|-----|----|--------|---------------------|-----|---|-----|-----|----|---------|---------------------|
| 531 | 7 | 0.5 | 360 | 14 | 039839 | 039839 influenza a | 604 | 7 | 0.5 | 415 | 4 | 095906 | 095906 homo sapien |
| 532 | 7 | 0.5 | 363 | 2 | 025617 | 025617 helicobacte | 605 | 7 | 0.5 | 415 | 4 | 09NRD5 | 09NRD5 homo sapien |
| 533 | 7 | 0.5 | 367 | 4 | 09P018 | 09P018 homo sapien | 606 | 7 | 0.5 | 416 | 10 | 09SVG6 | 09SVG6 arabidopsis |
| 534 | 7 | 0.5 | 368 | 10 | 09LIC1 | 09LIC1 arabidopsis | 607 | 7 | 0.5 | 417 | 2 | 09WZ24 | 09WZ24 thermotoga |
| 535 | 7 | 0.5 | 368 | 10 | 09S711 | 09S711 arabidopsis | 608 | 7 | 0.5 | 420 | 4 | 014979 | 014979 homo sapien |
| 536 | 7 | 0.5 | 369 | 2 | 09JW8 | 09JW8 neisseria m | 609 | 7 | 0.5 | 420 | 4 | 01NSR2 | 01NSR2 homo sapien |
| 537 | 7 | 0.5 | 370 | 11 | 035987 | 035987 rattus norv | 610 | 7 | 0.5 | 420 | 10 | 09SJF6 | 09SJF6 arabidopsis |
| 538 | 7 | 0.5 | 371 | 5 | 09VTS9 | 09VTS9 drosophila | 611 | 7 | 0.5 | 421 | 5 | 09U6S0 | 09U6S0 strongyloce |
| 539 | 7 | 0.5 | 372 | 2 | 074455 | 074455 synechocyst | 612 | 7 | 0.5 | 422 | 5 | 09W4E5 | 09W4E5 drosophila |
| 540 | 7 | 0.5 | 372 | 10 | 080888 | 080888 arabidopsis | 613 | 7 | 0.5 | 422 | 14 | 092503 | 092503 bombyx mori |
| 541 | 7 | 0.5 | 373 | 4 | 09UQ42 | 09UQ42 homo sapien | 614 | 7 | 0.5 | 423 | 10 | 09LQO8 | 09LQO8 arabidopsis |
| 542 | 7 | 0.5 | 373 | 4 | 09Y2Y7 | 09Y2Y7 homo sapien | 615 | 7 | 0.5 | 424 | 5 | 017469 | 017469 caenorhabdi |
| 543 | 7 | 0.5 | 373 | 5 | 026960 | 026960 trypanosoma | 616 | 7 | 0.5 | 424 | 10 | 09ZNR7 | 09ZNR7 arabidopsis |
| 544 | 7 | 0.5 | 373 | 10 | 09LR97 | 09LR97 arabidopsis | 617 | 7 | 0.5 | 426 | 5 | 09NF73 | 09NF73 drosophila |
| 545 | 7 | 0.5 | 374 | 5 | 09VVB5 | 09VVB5 drosophila | 618 | 7 | 0.5 | 427 | 3 | 071682 | 071682 aspergillus |
| 546 | 7 | 0.5 | 374 | 14 | 09WFI5 | 09WFI5 human adeno | 619 | 7 | 0.5 | 427 | 11 | 064321 | 064321 mus musculus |
| 547 | 7 | 0.5 | 376 | 2 | 09ZC09 | 09ZC09 streptomyce | 620 | 7 | 0.5 | 429 | 3 | 09P6R2 | 09P6R2 schizosacch |
| 548 | 7 | 0.5 | 377 | 11 | 09Z223 | 09Z223 cricetus | 621 | 7 | 0.5 | 429 | 10 | 09SAD0 | 09SAD0 arabidopsis |
| 549 | 7 | 0.5 | 378 | 5 | 09VRR9 | 09VRR9 drosophila | 622 | 7 | 0.5 | 431 | 1 | 027685 | 027685 methanobact |
| 550 | 7 | 0.5 | 379 | 5 | 025705 | 025705 plasmodium | 623 | 7 | 0.5 | 431 | 10 | 09FLP9 | 09FLP9 arabidopsis |
| 551 | 7 | 0.5 | 379 | 5 | 025706 | 025706 plasmodium | 624 | 7 | 0.5 | 433 | 2 | 09RTF6 | 09RTF6 deinococcus |
| 552 | 7 | 0.5 | 379 | 5 | 09U6C4 | 09U6C4 plasmodium | 625 | 7 | 0.5 | 434 | 10 | 065310 | 065310 arabidopsis |
| 553 | 7 | 0.5 | 379 | 10 | 09W8S5 | 09W8S5 arabidopsis | 626 | 7 | 0.5 | 434 | 10 | 09SBJ5 | 09SBJ5 arabidopsis |
| 554 | 7 | 0.5 | 380 | 5 | 026019 | 026019 plasmodium | 627 | 7 | 0.5 | 435 | 10 | 022281 | 022281 arabidopsis |
| 555 | 7 | 0.5 | 380 | 5 | 09XZ72 | 09XZ72 libellula p | 628 | 7 | 0.5 | 435 | 10 | 09FKN6 | 09FKN6 arabidopsis |
| 556 | 7 | 0.5 | 381 | 3 | 09P625 | 09P625 neurospora | 629 | 7 | 0.5 | 436 | 3 | 036020 | 036020 schizosacch |
| 557 | 7 | 0.5 | 381 | 4 | 09UQC5 | 09UQC5 homo sapien | 630 | 7 | 0.5 | 436 | 4 | 043440 | 043440 homo sapien |
| 558 | 7 | 0.5 | 382 | 5 | 09GQ81 | 09GQ81 drosophila | 631 | 7 | 0.5 | 438 | 11 | 055036 | 055036 cricetus |
| 559 | 7 | 0.5 | 382 | 8 | 099971 | 099971 porphyra pu | 632 | 7 | 0.5 | 438 | 11 | 09ER75 | 09ER75 mus musculus |
| 560 | 7 | 0.5 | 382 | 11 | 09P5W8 | 09P5W8 xenopus lae | 633 | 7 | 0.5 | 439 | 2 | 09L516 | 09L516 salmonella |
| 561 | 7 | 0.5 | 383 | 11 | 09WVM2 | 09WVM2 mus musculus | 634 | 7 | 0.5 | 439 | 5 | 09GTH6 | 09GTH6 caenorhabdi |
| 562 | 7 | 0.5 | 384 | 5 | 09XZ71 | 09XZ71 periplaneta | 635 | 7 | 0.5 | 440 | 4 | 09NV78 | 09NV78 homo sapien |
| 563 | 7 | 0.5 | 384 | 5 | 09V394 | 09V394 drosophila | 636 | 7 | 0.5 | 441 | 4 | 09NW01 | 09NW01 homo sapien |
| 564 | 7 | 0.5 | 386 | 3 | 09P7B0 | 09P7B0 schizosacch | 637 | 7 | 0.5 | 441 | 10 | 09STC9 | 09STC9 zinnia eleg |
| 565 | 7 | 0.5 | 386 | 3 | 044534 | 044534 aedes aegypt | 638 | 7 | 0.5 | 443 | 10 | 09ZRT26 | 09ZRT26 perilla fru |
| 566 | 7 | 0.5 | 386 | 13 | 091839 | 091839 xenopus lae | 639 | 7 | 0.5 | 443 | 13 | 0918W3 | 0918W3 gallus gall |
| 567 | 7 | 0.5 | 386 | 14 | 091MNA | 091MNA hydrangea m | 640 | 7 | 0.5 | 444 | 4 | 09UPW2 | 09UPW2 homo sapien |
| 568 | 7 | 0.5 | 387 | 2 | 091296 | 091296 pseudomonas | 641 | 7 | 0.5 | 445 | 5 | 09VN84 | 09VN84 drosophila |
| 569 | 7 | 0.5 | 388 | 5 | 09VVB3 | 09VVB3 drosophila | 642 | 7 | 0.5 | 445 | 10 | 09ZWC5 | 09ZWC5 arabidopsis |
| 570 | 7 | 0.5 | 388 | 10 | 048678 | 048678 arabidopsis | 643 | 7 | 0.5 | 445 | 10 | 09W1Y5 | 09W1Y5 arabidopsis |
| 571 | 7 | 0.5 | 388 | 13 | 09DF24 | 09DF24 xenopus lae | 644 | 7 | 0.5 | 445 | 11 | 060877 | 060877 mus musculus |
| 572 | 7 | 0.5 | 389 | 6 | 09GL07 | 09GL07 phoca vitul | 645 | 7 | 0.5 | 445 | 11 | 035371 | 035371 mus musculus |
| 573 | 7 | 0.5 | 391 | 5 | 09VTT2 | 09VTT2 drosophila | 646 | 7 | 0.5 | 447 | 4 | 09H931 | 09H931 homo sapien |
| 574 | 7 | 0.5 | 391 | 6 | 09GL06 | 09GL06 sus scrofa | 647 | 7 | 0.5 | 448 | 4 | 09Y2V1 | 09Y2V1 homo sapien |
| 575 | 7 | 0.5 | 391 | 10 | 09L127 | 09L127 oryza sativ | 648 | 7 | 0.5 | 448 | 6 | 09XS63 | 09XS63 equus caball |
| 576 | 7 | 0.5 | 392 | 4 | 013149 | 013149 homo sapien | 649 | 7 | 0.5 | 448 | 10 | 09STM4 | 09STM4 arabidopsis |
| 577 | 7 | 0.5 | 393 | 3 | 013701 | 013701 schizosacch | 650 | 7 | 0.5 | 449 | 4 | 09UE16 | 09UE16 homo sapien |
| 578 | 7 | 0.5 | 393 | 6 | 09GL16 | 09GL16 hippopotamu | 651 | 7 | 0.5 | 450 | 10 | 09SUC3 | 09SUC3 arabidopsis |
| 579 | 7 | 0.5 | 393 | 6 | 09GL15 | 09GL15 macrotus ca | 652 | 7 | 0.5 | 452 | 5 | 019175 | 019175 caenorhabdi |
| 580 | 7 | 0.5 | 393 | 6 | 09GL11 | 09GL11 nycticebus | 653 | 7 | 0.5 | 452 | 10 | 09SGX8 | 09SGX8 arabidopsis |
| 581 | 7 | 0.5 | 393 | 11 | 070278 | 070278 mus musculus | 654 | 7 | 0.5 | 454 | 10 | 004500 | 004500 arabidopsis |
| 582 | 7 | 0.5 | 394 | 1 | 09UYC0 | 09UYC0 pyrococcus | 655 | 7 | 0.5 | 454 | 11 | 09QY91 | 09QY91 mus musculus |
| 583 | 7 | 0.5 | 395 | 6 | 09GL12 | 09GL12 manis sp. a | 656 | 7 | 0.5 | 455 | 3 | 012080 | 012080 saccharomyc |
| 584 | 7 | 0.5 | 395 | 13 | 042481 | 042481 makaira nig | 657 | 7 | 0.5 | 455 | 11 | 062511 | 062511 mus musculus |
| 585 | 7 | 0.5 | 396 | 2 | 09F793 | 09F793 campylobact | 658 | 7 | 0.5 | 456 | 10 | 09SL95 | 09SL95 arabidopsis |
| 586 | 7 | 0.5 | 396 | 11 | 09Z1K1 | 09Z1K1 mus musculus | 659 | 7 | 0.5 | 457 | 4 | 09NSD2 | 09NSD2 homo sapien |
| 587 | 7 | 0.5 | 397 | 5 | 09VVB4 | 09VVB4 drosophila | 660 | 7 | 0.5 | 457 | 4 | 09HAK6 | 09HAK6 homo sapien |
| 588 | 7 | 0.5 | 398 | 2 | 052121 | 052121 escherichia | 661 | 7 | 0.5 | 460 | 4 | 09ULW6 | 09ULW6 homo sapien |
| 589 | 7 | 0.5 | 399 | 4 | 09NFB0 | 09NFB0 homo sapien | 662 | 7 | 0.5 | 461 | 2 | 09I6E1 | 09I6E1 pseudomonas |
| 590 | 7 | 0.5 | 399 | 10 | 09LW33 | 09LW33 arabidopsis | 663 | 7 | 0.5 | 461 | 10 | 09ZRG9 | 09ZRG9 oryza sativ |
| 591 | 7 | 0.5 | 400 | 5 | 027934 | 027934 drosophila | 664 | 7 | 0.5 | 461 | 10 | 09FGZ7 | 09FGZ7 arabidopsis |
| 592 | 7 | 0.5 | 400 | 10 | 09ZRM3 | 09ZRM3 antirrhinum | 665 | 7 | 0.5 | 463 | 2 | 09RNM3 | 09RNM3 zymomonas m |
| 593 | 7 | 0.5 | 405 | 4 | 013215 | 013215 homo sapien | 666 | 7 | 0.5 | 464 | 4 | 043159 | 043159 homo sapien |
| 594 | 7 | 0.5 | 405 | 5 | 026268 | 026268 aplysia cal | 667 | 7 | 0.5 | 466 | 5 | 093712 | 093712 caenorhabdi |
| 595 | 7 | 0.5 | 405 | 11 | 09WV69 | 09WV69 mus musculus | 668 | 7 | 0.5 | 469 | 1 | 072021 | 072021 methanosarc |
| 596 | 7 | 0.5 | 407 | 5 | 09TX81 | 09TX81 plasmodium | 669 | 7 | 0.5 | 471 | 2 | 09HWN7 | 09HWN7 pseudomonas |
| 597 | 7 | 0.5 | 407 | 10 | 049747 | 049747 arabidopsis | 670 | 7 | 0.5 | 471 | 4 | 09H0V2 | 09H0V2 homo sapien |
| 598 | 7 | 0.5 | 409 | 10 | 09SGC6 | 09SGC6 arabidopsis | 671 | 7 | 0.5 | 471 | 10 | 09LXJ8 | 09LXJ8 arabidopsis |
| 599 | 7 | 0.5 | 409 | 10 | 09LN21 | 09LN21 arabidopsis | 672 | 7 | 0.5 | 472 | 1 | 09VZ24 | 09VZ24 methanosarc |
| 600 | 7 | 0.5 | 412 | 4 | 013558 | 013558 homo sapien | 673 | 7 | 0.5 | 473 | 13 | 073840 | 073840 gallus gall |
| 601 | 7 | 0.5 | 412 | 4 | 09ULX3 | 09ULX3 homo sapien | 674 | 7 | 0.5 | 474 | 4 | 09UK66 | 09UK66 homo sapien |
| 602 | 7 | 0.5 | 412 | 10 | 09FZC9 | 09FZC9 arabidopsis | 675 | 7 | 0.5 | 475 | 1 | 058382 | 058382 pyrococcus |
| 603 | 7 | 0.5 | 414 | 3 | 09P5H7 | 09P5H7 neurospora | 676 | 7 | 0.5 | 476 | 4 | 09UJU8 | 09UJU8 homo sapien |

| | | | | | | | | | | | | | |
|-----|---|-----|-----|----|---------|---------------------|-----|---|-----|-----|----|--------|---------------------|
| 385 | 7 | 0.5 | 270 | 5 | 096536 | O96536 strongyloce | 458 | 7 | 0.5 | 316 | 10 | 09FX02 | O9fx02 arabidopsis |
| 386 | 7 | 0.5 | 271 | 5 | 09TZ06 | Q9tzq6 strongyloce | 459 | 7 | 0.5 | 316 | 11 | Q62254 | Q62254 mus musculus |
| 387 | 7 | 0.5 | 270 | 2 | Q9KUK5 | Q9kuk5 vibrio chol | 460 | 7 | 0.5 | 316 | 11 | P70113 | P70113 mesocricetu |
| 388 | 7 | 0.5 | 271 | 10 | Q9SV06 | Q9sv06 arabidopsis | 461 | 7 | 0.5 | 320 | 4 | Q9H4R0 | Q9h4r0 homo sapien |
| 389 | 7 | 0.5 | 273 | 10 | Q9FHB2 | Q9fhb2 arabidopsis | 462 | 7 | 0.5 | 321 | 6 | P79119 | P79119 bos taurus |
| 390 | 7 | 0.5 | 275 | 13 | Q13090 | Q13090 pleurodeles | 463 | 7 | 0.5 | 322 | 4 | Q99645 | Q99645 homo sapien |
| 391 | 7 | 0.5 | 276 | 2 | Q9F733 | Q9f733 chlorobium | 464 | 7 | 0.5 | 323 | 2 | P95185 | P95185 mycobacteri |
| 392 | 7 | 0.5 | 276 | 13 | P79780 | P79780 gallus gall | 465 | 7 | 0.5 | 323 | 13 | Q73907 | Q73907 gallus gall |
| 393 | 7 | 0.5 | 279 | 10 | Q9FHY1 | Q9fhy1 arabidopsis | 466 | 7 | 0.5 | 324 | 2 | Q92834 | Q92834 chlamydia p |
| 394 | 7 | 0.5 | 280 | 4 | Q9NVW5 | Q9nvw5 homo sapien | 467 | 7 | 0.5 | 324 | 3 | Q9HEF2 | Q9hef2 neurospora |
| 395 | 7 | 0.5 | 283 | 13 | Q9I939 | Q9i939 misgurnus a | 468 | 7 | 0.5 | 326 | 4 | Q9HAJ4 | Q9haj4 homo sapien |
| 396 | 7 | 0.5 | 284 | 4 | Q9H215 | Q9h215 homo sapien | 469 | 7 | 0.5 | 326 | 5 | Q9VPX6 | Q9vpx6 drosophila |
| 397 | 7 | 0.5 | 285 | 11 | Q9Z2D8 | Q9z2d8 mus musculus | 470 | 7 | 0.5 | 328 | 10 | Q9ZVX7 | Q9zvx7 arabidopsis |
| 398 | 7 | 0.5 | 286 | 2 | Q50196 | Q50196 mycobacteri | 471 | 7 | 0.5 | 328 | 13 | Q9PRJ2 | Q9prj2 brachydanio |
| 399 | 7 | 0.5 | 286 | 14 | Q9DH21 | Q9dh21 tt virus. v | 472 | 7 | 0.5 | 329 | 2 | O84436 | O84436 chlamydia t |
| 400 | 7 | 0.5 | 288 | 5 | O17375 | O17375 caenorhabdi | 473 | 7 | 0.5 | 329 | 4 | Q9NVW8 | Q9nvw8 homo sapien |
| 401 | 7 | 0.5 | 289 | 11 | Q9JK23 | Q9jk23 mus musculus | 474 | 7 | 0.5 | 329 | 5 | Q9NFV9 | Q9nfv9 plasmodium |
| 402 | 7 | 0.5 | 290 | 14 | Q9Q8V0 | Q9q8v0 shope fibro | 475 | 7 | 0.5 | 329 | 10 | Q9FH65 | Q9fh65 arabidopsis |
| 403 | 7 | 0.5 | 291 | 2 | Q9S0F2 | Q9s0f2 borrelia bu | 476 | 7 | 0.5 | 330 | 6 | Q9GLY8 | Q9gly8 macaca fasc |
| 404 | 7 | 0.5 | 291 | 2 | Q9S0C3 | Q9s0c3 borrelia bu | 477 | 7 | 0.5 | 331 | 2 | Q9SLJ8 | Q9slj8 streptococc |
| 405 | 7 | 0.5 | 291 | 2 | Q9S071 | Q9s071 borrelia bu | 478 | 7 | 0.5 | 331 | 2 | Q9FDD3 | Q9fdt3 brucella ab |
| 406 | 7 | 0.5 | 291 | 2 | Q9S046 | Q9s046 borrelia bu | 479 | 7 | 0.5 | 331 | 3 | Q9V7T2 | Q9v7t2 schizosacch |
| 407 | 7 | 0.5 | 291 | 2 | Q9RZ25 | Q9rzz5 borrelia bu | 480 | 7 | 0.5 | 331 | 4 | Q9HD90 | Q9hd90 homo sapien |
| 408 | 7 | 0.5 | 291 | 2 | Q9R331 | Q9r331 borrelia bu | 481 | 7 | 0.5 | 331 | 10 | Q9FX56 | Q9fx56 arabidopsis |
| 409 | 7 | 0.5 | 291 | 4 | Q9S983 | Q9s983 homo sapien | 482 | 7 | 0.5 | 332 | 2 | Q9PJW3 | Q9pjw3 chlamydia m |
| 410 | 7 | 0.5 | 291 | 4 | Q9NUL5 | Q9nul5 homo sapien | 483 | 7 | 0.5 | 332 | 10 | Q9LQW2 | Q9lqw2 arabidopsis |
| 411 | 7 | 0.5 | 292 | 2 | O44793 | O44793 borrelia bu | 484 | 7 | 0.5 | 335 | 10 | Q9S9Q6 | Q9s9q6 arabidopsis |
| 412 | 7 | 0.5 | 292 | 2 | Q9X217 | Q9x217 thermotoga | 485 | 7 | 0.5 | 337 | 5 | Q9Y156 | Q9y156 drosophila |
| 413 | 7 | 0.5 | 292 | 2 | Q9S088 | Q9s088 borrelia bu | 486 | 7 | 0.5 | 337 | 14 | Q9DM07 | Q9dm07 plutella xy |
| 414 | 7 | 0.5 | 292 | 2 | Q9L9N7 | Q9l9n7 borrelia bu | 487 | 7 | 0.5 | 339 | 10 | Q9LQW4 | Q9lqw4 arabidopsis |
| 415 | 7 | 0.5 | 292 | 10 | Q9Z730 | Q9z730 arabidopsis | 488 | 7 | 0.5 | 339 | 11 | Q61149 | Q61149 mus musculus |
| 416 | 7 | 0.5 | 292 | 10 | Q9Z090 | Q9zu90 arabidopsis | 489 | 7 | 0.5 | 340 | 2 | Q9KNS0 | Q9kns0 vibrio chol |
| 417 | 7 | 0.5 | 293 | 4 | Q9UDC3 | Q9udc3 homo sapien | 490 | 7 | 0.5 | 340 | 4 | Q9NXZ4 | Q9nxz4 homo sapien |
| 418 | 7 | 0.5 | 293 | 10 | Q9L179 | Q9l179 arabidopsis | 491 | 7 | 0.5 | 340 | 5 | Q9U7C9 | Q9u7c9 dictyosteli |
| 419 | 7 | 0.5 | 293 | 10 | Q9FJ35 | Q9fj35 arabidopsis | 492 | 7 | 0.5 | 340 | 10 | Q9LFS1 | Q9lfs1 arabidopsis |
| 420 | 7 | 0.5 | 293 | 13 | Q9I8M8 | Q9i8md petromyzon | 493 | 7 | 0.5 | 341 | 3 | O74509 | O74509 schizosacch |
| 421 | 7 | 0.5 | 293 | 14 | Q91824 | Q91824 coho salmon | 494 | 7 | 0.5 | 342 | 5 | Q9NGS6 | Q9ngs6 acropora mi |
| 422 | 7 | 0.5 | 294 | 14 | Q9QF47 | Q9qf47 avian pneum | 495 | 7 | 0.5 | 343 | 10 | O64878 | O64878 arabidopsis |
| 423 | 7 | 0.5 | 295 | 10 | Q9ZW81 | Q9zwm1 arabidopsis | 496 | 7 | 0.5 | 344 | 11 | Q9ER45 | Q9er45 mus musculus |
| 424 | 7 | 0.5 | 295 | 10 | Q9LSZ8 | Q9lsz8 arabidopsis | 497 | 7 | 0.5 | 346 | 5 | Q9U0G0 | Q9u0g0 plasmodium |
| 425 | 7 | 0.5 | 296 | 2 | Q9Z8S5 | Q9z8s5 chlamydia p | 498 | 7 | 0.5 | 347 | 4 | Q9UKP3 | Q9ukp3 homo sapien |
| 426 | 7 | 0.5 | 297 | 5 | Q9VEM3 | Q9vem3 drosophila | 499 | 7 | 0.5 | 347 | 4 | Q9HAF9 | Q9haf9 homo sapien |
| 427 | 7 | 0.5 | 298 | 1 | Q9EM205 | Q9em205 pyrococcus | 500 | 7 | 0.5 | 347 | 5 | O61872 | O61872 caenorhabdi |
| 428 | 7 | 0.5 | 298 | 3 | Q959842 | Q959842 penicillium | 501 | 7 | 0.5 | 347 | 13 | Q9DE43 | Q9de43 brachydanio |
| 429 | 7 | 0.5 | 299 | 1 | Q9UYE4 | Q9uye4 pyrococcus | 502 | 7 | 0.5 | 348 | 10 | Q9LU06 | Q9lu06 arabidopsis |
| 430 | 7 | 0.5 | 299 | 5 | Q9VWL3 | Q9vwl3 drosophila | 503 | 7 | 0.5 | 348 | 10 | Q9FLX8 | Q9flx8 arabidopsis |
| 431 | 7 | 0.5 | 299 | 10 | O82801 | O82801 arabidopsis | 504 | 7 | 0.5 | 349 | 10 | Q9LU07 | Q9lu07 arabidopsis |
| 432 | 7 | 0.5 | 299 | 10 | O48722 | O48722 arabidopsis | 505 | 7 | 0.5 | 349 | 13 | Q9YHY7 | Q9yhy7 gallus gall |
| 433 | 7 | 0.5 | 299 | 11 | Q9QX01 | Q9qyu1 rattus ratt | 506 | 7 | 0.5 | 350 | 4 | Q9S455 | Q9s455 homo sapien |
| 434 | 7 | 0.5 | 300 | 6 | O46383 | O46383 mus musculus | 507 | 7 | 0.5 | 350 | 5 | O76720 | O76720 caenorhabdi |
| 435 | 7 | 0.5 | 301 | 2 | Q9RGF0 | Q9rgf0 campylobact | 508 | 7 | 0.5 | 350 | 5 | Q9NE21 | Q9ne21 leishmania |
| 436 | 7 | 0.5 | 301 | 2 | O45333 | O45333 caenorhabdi | 509 | 7 | 0.5 | 350 | 10 | Q9SMQ4 | Q9smq4 arabidopsis |
| 437 | 7 | 0.5 | 301 | 11 | Q9Z130 | Q9z130 mus musculus | 510 | 7 | 0.5 | 350 | 13 | O42202 | O42202 brachydanio |
| 438 | 7 | 0.5 | 302 | 10 | Q9FX68 | Q9fx68 arabidopsis | 511 | 7 | 0.5 | 351 | 4 | Q9ULR0 | Q9ulr0 homo sapien |
| 439 | 7 | 0.5 | 302 | 10 | Q9FN16 | Q9fn16 arabidopsis | 512 | 7 | 0.5 | 351 | 10 | Q9ZV08 | Q9zvu8 arabidopsis |
| 440 | 7 | 0.5 | 302 | 9 | Q9ZX05 | Q9zx05 mycobacteri | 513 | 7 | 0.5 | 351 | 10 | Q9M4F3 | Q9m4f3 brassica na |
| 441 | 7 | 0.5 | 303 | 10 | Q9M0H2 | Q9m0h2 arabidopsis | 514 | 7 | 0.5 | 351 | 10 | Q9FX55 | Q9fx55 arabidopsis |
| 442 | 7 | 0.5 | 303 | 10 | Q9LYP4 | Q9lyp4 arabidopsis | 515 | 7 | 0.5 | 352 | 11 | Q9R192 | Q9r192 rattus norv |
| 443 | 7 | 0.5 | 304 | 10 | Q9FRA5 | Q9fra5 oryza sativ | 516 | 7 | 0.5 | 353 | 4 | Q9UN41 | Q9un41 homo sapien |
| 444 | 7 | 0.5 | 306 | 2 | O67883 | O67883 aquifex aeo | 517 | 7 | 0.5 | 353 | 10 | Q9SAL5 | Q9sal5 arabidopsis |
| 445 | 7 | 0.5 | 306 | 2 | O17377 | O17377 caenorhabdi | 518 | 7 | 0.5 | 353 | 10 | Q9LVN5 | Q9lvn5 arabidopsis |
| 446 | 7 | 0.5 | 307 | 2 | O52061 | O52061 bacillus me | 519 | 7 | 0.5 | 354 | 5 | Q25995 | Q25995 plasmodium |
| 447 | 7 | 0.5 | 308 | 4 | Q9HAY2 | Q9hay2 homo sapien | 520 | 7 | 0.5 | 354 | 11 | Q9FK12 | Q9fk12 arabidopsis |
| 448 | 7 | 0.5 | 310 | 10 | Q9L7B9 | Q9l7b9 arabidopsis | 521 | 7 | 0.5 | 354 | 11 | O62509 | O62509 mus musculus |
| 449 | 7 | 0.5 | 314 | 1 | Q9HJW0 | Q9hjm0 thermoplasm | 522 | 7 | 0.5 | 355 | 1 | Q9UX70 | Q9ux70 sulfolobus |
| 450 | 7 | 0.5 | 314 | 5 | Q20146 | Q20146 caenorhabdi | 523 | 7 | 0.5 | 355 | 2 | Q9L1A1 | Q9l1a1 streptomyce |
| 451 | 7 | 0.5 | 314 | 11 | Q9ESH2 | Q9esh2 rattus norv | 524 | 7 | 0.5 | 356 | 4 | Q9UEC8 | Q9uec8 homo sapien |
| 452 | 7 | 0.5 | 315 | 4 | Q9NUD4 | Q9nud4 homo sapien | 525 | 7 | 0.5 | 356 | 11 | Q9R193 | Q9r193 rattus norv |
| 453 | 7 | 0.5 | 315 | 5 | O44605 | O44605 caenorhabdi | 526 | 7 | 0.5 | 357 | 2 | O83906 | O83906 treponema p |
| 454 | 7 | 0.5 | 316 | 2 | P75300 | P75300 mycoplasma | 527 | 7 | 0.5 | 357 | 4 | Q9UN42 | Q9un42 homo sapien |
| 455 | 7 | 0.5 | 316 | 10 | Q9MAA2 | Q9maa2 arabidopsis | 528 | 7 | 0.5 | 359 | 5 | Q9Y0C6 | Q9y0c6 trichinella |
| 456 | 7 | 0.5 | 316 | 10 | Q9MAA2 | Q9maa2 arabidopsis | 529 | 7 | 0.5 | 359 | 5 | Q9XV41 | Q9xv41 caenorhabdi |
| 457 | 7 | 0.5 | 316 | 10 | Q9MAA2 | Q9maa2 arabidopsis | 530 | 7 | 0.5 | 360 | 10 | Q9FFP9 | Q9ffp9 arabidopsis |

| | | | | | | | | | | | | | |
|-----|---|-----|-----|----|---------|---------------------|-----|---|-----|-----|----|--------|---------------------|
| 239 | 7 | 0.5 | 199 | 14 | Q9W9X3 | Q9w9x3 porcine ade | 312 | 7 | 0.5 | 235 | 5 | 096535 | 096535 strongyloce |
| 240 | 7 | 0.5 | 200 | 4 | Q15347 | Q15347 homo sapien | 313 | 7 | 0.5 | 235 | 10 | Q9L122 | Q9l122 oryza sativ |
| 241 | 7 | 0.5 | 200 | 10 | Q23294 | Q23294 arabisdopsis | 314 | 7 | 0.5 | 236 | 2 | Q9R527 | Q9r527 bacteroides |
| 242 | 7 | 0.5 | 201 | 11 | Q54879 | Q54879 mus musculu | 315 | 7 | 0.5 | 236 | 5 | Q24872 | Q24872 echinometra |
| 243 | 7 | 0.5 | 201 | 4 | Q95556 | Q95556 homo sapien | 316 | 7 | 0.5 | 236 | 5 | Q24874 | Q24874 echinometra |
| 244 | 7 | 0.5 | 202 | 2 | Q9F027 | Q9f027 bradyrhizob | 317 | 7 | 0.5 | 236 | 5 | Q24875 | Q24875 echinometra |
| 245 | 7 | 0.5 | 203 | 3 | Q9P7J2 | Q9p7j2 schizosacch | 318 | 7 | 0.5 | 236 | 5 | Q96538 | Q96538 hemicrotrot |
| 246 | 7 | 0.5 | 204 | 10 | Q92954 | Q92954 arabisdopsis | 319 | 7 | 0.5 | 236 | 5 | Q9T2Q8 | Q9t2q8 strongyloce |
| 247 | 7 | 0.5 | 204 | 10 | Q43386 | Q43386 arabisdopsis | 320 | 7 | 0.5 | 236 | 5 | Q9T2Q7 | Q9t2q7 strongyloce |
| 248 | 7 | 0.5 | 205 | 2 | Q31812 | Q31812 bacillus su | 321 | 7 | 0.5 | 236 | 5 | Q9T2Q5 | Q9t2q5 hemicrotrot |
| 249 | 7 | 0.5 | 206 | 2 | Q96510 | Q96510 borrelia af | 322 | 7 | 0.5 | 236 | 10 | Q9L1W6 | Q9l1w6 oryza sativ |
| 250 | 7 | 0.5 | 206 | 2 | Q9L6N7 | Q9l6n7 salmonella | 323 | 7 | 0.5 | 238 | 4 | Q9H032 | Q9h032 homo sapien |
| 251 | 7 | 0.5 | 208 | 2 | Q57530 | Q57530 agrobacteri | 324 | 7 | 0.5 | 238 | 10 | Q9FNL9 | Q9fnl9 arabisdopsis |
| 252 | 7 | 0.5 | 208 | 10 | Q9T089 | Q9t089 arabisdopsis | 325 | 7 | 0.5 | 238 | 14 | Q9WFX9 | Q9wfy9 human herpe |
| 253 | 7 | 0.5 | 210 | 4 | Q95054 | Q95054 homo sapien | 326 | 7 | 0.5 | 238 | 14 | Q9WFX8 | Q9wfy8 human herpe |
| 254 | 7 | 0.5 | 211 | 11 | Q9EQD5 | Q9eqd5 mus musculu | 327 | 7 | 0.5 | 239 | 6 | Q9W891 | Q9w891 human herpe |
| 255 | 7 | 0.5 | 211 | 2 | Q9Z357 | Q9z357 borrelia va | 328 | 7 | 0.5 | 239 | 6 | Q9XSK7 | Q9xsk7 bos taurus |
| 256 | 7 | 0.5 | 211 | 5 | Q9NAA6 | Q9naa6 caenorhabdi | 329 | 7 | 0.5 | 239 | 14 | Q9W8W3 | Q9w8w3 human herpe |
| 257 | 7 | 0.5 | 211 | 11 | Q9JHX0 | Q9jhx0 rattus norv | 330 | 7 | 0.5 | 241 | 5 | Q24907 | Q24907 echinometra |
| 258 | 7 | 0.5 | 213 | 4 | Q9UJG1 | Q9ujg1 homo sapien | 331 | 7 | 0.5 | 241 | 5 | Q24906 | Q24906 echinometra |
| 259 | 7 | 0.5 | 213 | 10 | Q43600 | Q43600 oryza sativ | 332 | 7 | 0.5 | 241 | 5 | Q24905 | Q24905 echinometra |
| 260 | 7 | 0.5 | 213 | 10 | Q9LPM6 | Q9lpw6 arabisdopsis | 333 | 7 | 0.5 | 241 | 10 | Q9ZQ05 | Q9zqg5 arabisdopsis |
| 261 | 7 | 0.5 | 214 | 1 | Q23215 | Q23215 archaeoglob | 334 | 7 | 0.5 | 242 | 5 | Q24909 | Q24909 echinometra |
| 262 | 7 | 0.5 | 214 | 3 | Q42932 | Q42932 schizosacch | 335 | 7 | 0.5 | 242 | 10 | Q9XE71 | Q9xe71 sorghum bic |
| 263 | 7 | 0.5 | 214 | 10 | Q49780 | Q49780 arabisdopsis | 336 | 7 | 0.5 | 242 | 10 | Q9LXD4 | Q9ldy4 oryza sativ |
| 264 | 7 | 0.5 | 214 | 13 | Q9PEUK9 | Q9pek9 gallus gall | 337 | 7 | 0.5 | 243 | 4 | Q9NKK9 | Q9nkk9 homo sapien |
| 265 | 7 | 0.5 | 215 | 2 | Q92EJ9 | Q9zef9 borrelia af | 338 | 7 | 0.5 | 243 | 4 | Q9H1T9 | Q9hit9 homo sapien |
| 266 | 7 | 0.5 | 215 | 2 | P70810 | P70810 borrelia af | 339 | 7 | 0.5 | 245 | 5 | Q24908 | Q24908 echinometra |
| 267 | 7 | 0.5 | 215 | 10 | Q65415 | Q65415 arabisdopsis | 340 | 7 | 0.5 | 245 | 10 | Q9FVE6 | Q9fve6 arabisdopsis |
| 268 | 7 | 0.5 | 215 | 11 | Q88611 | Q88611 spalax leuc | 341 | 7 | 0.5 | 247 | 5 | Q9W3H4 | Q9w3h4 drosophila |
| 269 | 7 | 0.5 | 215 | 13 | Q9YH06 | Q9yh06 gallus gall | 342 | 7 | 0.5 | 247 | 10 | Q81766 | Q81766 arabisdopsis |
| 270 | 7 | 0.5 | 215 | 14 | Q9QTF1 | Q9qtf1 unidentified | 343 | 7 | 0.5 | 248 | 5 | Q25676 | Q25676 plasmodium |
| 271 | 7 | 0.5 | 216 | 6 | Q9N213 | Q9n213 macaca sylv | 344 | 7 | 0.5 | 248 | 10 | Q80892 | Q80892 arabisdopsis |
| 272 | 7 | 0.5 | 216 | 6 | Q9N209 | Q9n209 macaca sylv | 345 | 7 | 0.5 | 248 | 10 | Q80893 | Q80893 arabisdopsis |
| 273 | 7 | 0.5 | 216 | 6 | Q9N207 | Q9n207 callithrix | 346 | 7 | 0.5 | 249 | 5 | Q9UIN3 | Q9uin3 cryptospori |
| 274 | 7 | 0.5 | 216 | 10 | Q9SLG7 | Q9slg7 arabisdopsis | 347 | 7 | 0.5 | 249 | 9 | Q80305 | Q80305 bacterioph |
| 275 | 7 | 0.5 | 216 | 11 | Q08904 | Q08904 mus musculu | 348 | 7 | 0.5 | 249 | 10 | Q22952 | Q22952 arabisdopsis |
| 276 | 7 | 0.5 | 217 | 3 | Q9HEX7 | Q9hex7 pneumocysti | 349 | 7 | 0.5 | 250 | 2 | Q9ZDU4 | Q9zdu4 rickettsia |
| 277 | 7 | 0.5 | 217 | 4 | Q9H7Y2 | Q9h7y2 homo sapien | 350 | 7 | 0.5 | 250 | 2 | Q9JMP7 | Q9jmp7 wolbachia s |
| 278 | 7 | 0.5 | 217 | 5 | Q97461 | Q97461 strongyloce | 351 | 7 | 0.5 | 250 | 5 | Q62387 | Q62387 caenorhabdi |
| 279 | 7 | 0.5 | 217 | 10 | Q9LXL1 | Q9lyl1 arabisdopsis | 352 | 7 | 0.5 | 250 | 5 | Q24917 | Q24917 echinometra |
| 280 | 7 | 0.5 | 217 | 13 | Q9IB51 | Q9ib51 oncorhynch | 353 | 7 | 0.5 | 250 | 11 | Q88795 | Q88795 mus musculu |
| 281 | 7 | 0.5 | 219 | 5 | Q96537 | Q96537 allocentrot | 354 | 7 | 0.5 | 251 | 5 | Q24916 | Q24916 echinometra |
| 282 | 7 | 0.5 | 219 | 5 | Q9TYD4 | Q9tyd4 allocentrot | 355 | 7 | 0.5 | 251 | 13 | Q9VI93 | Q9vi93 gallus gall |
| 283 | 7 | 0.5 | 221 | 10 | Q9ZVR9 | Q9zvr9 arabisdopsis | 356 | 7 | 0.5 | 254 | 10 | Q48767 | Q48767 arabisdopsis |
| 284 | 7 | 0.5 | 223 | 4 | Q9NQF9 | Q9nqf9 homo sapien | 357 | 7 | 0.5 | 255 | 4 | Q9NWB0 | Q9nwb0 homo sapien |
| 285 | 7 | 0.5 | 223 | 10 | Q9W2Y1 | Q9w2y1 arabisdopsis | 358 | 7 | 0.5 | 255 | 4 | P95273 | P95273 mycobacteri |
| 286 | 7 | 0.5 | 224 | 2 | Q55322 | Q55322 synechococc | 359 | 7 | 0.5 | 256 | 10 | Q9FSJ4 | Q9fsj4 oryza sativ |
| 287 | 7 | 0.5 | 224 | 4 | Q94797 | Q94797 homo sapien | 360 | 7 | 0.5 | 257 | 4 | Q9P1V9 | Q9p1v9 homo sapien |
| 288 | 7 | 0.5 | 228 | 1 | Q29479 | Q29479 archaeoglob | 361 | 7 | 0.5 | 257 | 10 | Q22238 | Q22238 arabisdopsis |
| 289 | 7 | 0.5 | 228 | 4 | Q9P0A8 | Q9p0a8 homo sapien | 362 | 7 | 0.5 | 259 | 2 | Q9Z9V7 | Q9z9v7 bacillus ha |
| 290 | 7 | 0.5 | 228 | 5 | Q96534 | Q96534 strongyloce | 363 | 7 | 0.5 | 259 | 2 | Q9KGA5 | Q9kga5 bacillus ha |
| 291 | 7 | 0.5 | 229 | 10 | Q24480 | Q24480 allium porr | 364 | 7 | 0.5 | 259 | 13 | Q91964 | Q91964 gallus gall |
| 292 | 7 | 0.5 | 230 | 3 | Q9URZ1 | Q9urz1 schizosacch | 365 | 7 | 0.5 | 260 | 5 | Q9XUN9 | Q9xun9 caenorhabdi |
| 293 | 7 | 0.5 | 230 | 5 | Q24870 | Q24870 echinometra | 366 | 7 | 0.5 | 260 | 11 | P97822 | P97822 mus musculu |
| 294 | 7 | 0.5 | 232 | 1 | Q28269 | Q28269 archaeoglob | 367 | 7 | 0.5 | 261 | 5 | Q9NF82 | Q9nf82 leishmania |
| 295 | 7 | 0.5 | 232 | 5 | Q24871 | Q24871 echinometra | 368 | 7 | 0.5 | 261 | 10 | Q9SK91 | Q9sk91 arabisdopsis |
| 296 | 7 | 0.5 | 232 | 10 | Q9ST20 | Q9st20 brassica ca | 369 | 7 | 0.5 | 262 | 2 | Q91AHL | Q91ahl pseudomonas |
| 297 | 7 | 0.5 | 233 | 2 | Q9HTY5 | Q9hty5 pseudomonas | 370 | 7 | 0.5 | 262 | 3 | Q74515 | Q74515 schizosacch |
| 298 | 7 | 0.5 | 233 | 5 | Q44377 | Q44377 arabia pun | 371 | 7 | 0.5 | 262 | 4 | Q9ULM8 | Q9ulm8 homo sapien |
| 299 | 7 | 0.5 | 233 | 5 | Q46361 | Q46361 arabia pun | 372 | 7 | 0.5 | 262 | 14 | Q92547 | Q92547 epizootic h |
| 300 | 7 | 0.5 | 234 | 6 | Q9T530 | Q9t530 oryctolag | 373 | 7 | 0.5 | 264 | 10 | Q04470 | Q04470 arabisdopsis |
| 301 | 7 | 0.5 | 235 | 2 | Q9PNC3 | Q9pnc3 campylobact | 374 | 7 | 0.5 | 265 | 2 | Q9KCA9 | Q9kca9 bacillus ha |
| 302 | 7 | 0.5 | 235 | 5 | Q44373 | Q44373 arabia lix | 375 | 7 | 0.5 | 265 | 10 | Q9M2Q7 | Q9m2q7 arabisdopsis |
| 303 | 7 | 0.5 | 235 | 5 | Q44374 | Q44374 arabia lix | 376 | 7 | 0.5 | 265 | 13 | Q98916 | Q98916 gallus gall |
| 304 | 7 | 0.5 | 235 | 5 | Q44375 | Q44375 arabia lix | 377 | 7 | 0.5 | 266 | 10 | Q9FZ93 | Q9fz93 arabisdopsis |
| 305 | 7 | 0.5 | 235 | 5 | Q44376 | Q44376 arabia lix | 378 | 7 | 0.5 | 266 | 10 | Q9FP61 | Q9fp61 oryza sativ |
| 306 | 7 | 0.5 | 235 | 5 | Q44378 | Q44378 arabia inc | 379 | 7 | 0.5 | 267 | 2 | Q9HWT0 | Q9hyt0 pseudomonas |
| 307 | 7 | 0.5 | 235 | 5 | Q44379 | Q44379 arabia inc | 380 | 7 | 0.5 | 267 | 5 | Q61076 | Q61076 plasmodium |
| 308 | 7 | 0.5 | 235 | 5 | Q46355 | Q46355 arabia inc | 381 | 7 | 0.5 | 267 | 10 | Q9SJ10 | Q9sj10 arabisdopsis |
| 309 | 7 | 0.5 | 235 | 5 | Q46366 | Q46366 arabia lix | 382 | 7 | 0.5 | 267 | 10 | Q9M9U2 | Q9m9u2 arabisdopsis |
| 310 | 7 | 0.5 | 235 | 5 | Q24873 | Q24873 echinometra | 383 | 7 | 0.5 | 269 | 5 | Q20892 | Q20892 caenorhabdi |
| 311 | 7 | 0.5 | 235 | 5 | Q46358 | Q46358 arabia duf | 384 | 7 | 0.5 | 270 | 2 | Q30448 | Q30448 bordetella |

| | | | | | | | | | | | | | |
|-----|---|-----|------|----|--------|---------------------|-----|---|-----|-----|----|--------|---------------------|
| 93 | 8 | 0.6 | 1082 | 11 | O55201 | 055201 mus musculus | 166 | 7 | 0.5 | 131 | 10 | P93015 | P93015 arabidopsis |
| 94 | 8 | 0.6 | 1095 | 4 | Q9HCG6 | Q9hcg6 homo sapien | 167 | 7 | 0.5 | 131 | 10 | Q9LKU5 | Q9liku5 arabidopsis |
| 95 | 8 | 0.6 | 1132 | 4 | Q9UF81 | Q9uf81 homo sapien | 168 | 7 | 0.5 | 134 | 2 | Q9RUW2 | Q9ruw2 deinococcus |
| 96 | 8 | 0.6 | 1173 | 11 | Q63624 | Q63624 rattus norv | 169 | 7 | 0.5 | 135 | 5 | Q9GT85 | Q9gt85 plasmodium |
| 97 | 8 | 0.6 | 1191 | 5 | Q9XYV3 | Q9xyv3 dictyosteli | 170 | 7 | 0.5 | 136 | 5 | Q94171 | Q94171 caenorhabdi |
| 98 | 8 | 0.6 | 1196 | 5 | Q23866 | Q23866 dictyosteli | 171 | 7 | 0.5 | 136 | 5 | Q9NH18 | Q9nh18 plasmodium |
| 99 | 8 | 0.6 | 1218 | 5 | Q9V646 | Q9v646 drosophila | 172 | 7 | 0.5 | 138 | 5 | Q20867 | Q20867 caenorhabdi |
| 100 | 8 | 0.6 | 1220 | 4 | Q9UMN7 | Q9umn7 homo sapien | 173 | 7 | 0.5 | 139 | 5 | O61947 | O61947 caenorhabdi |
| 101 | 8 | 0.6 | 1231 | 5 | O18017 | O18017 caenorhabdi | 174 | 7 | 0.5 | 140 | 10 | Q9SHZ7 | Q9shz7 arabidopsis |
| 102 | 8 | 0.6 | 1235 | 4 | Q9UL54 | Q9ul54 homo sapien | 175 | 7 | 0.5 | 141 | 9 | Q9MCJ2 | Q9mcj2 streptococc |
| 103 | 8 | 0.6 | 1235 | 11 | Q9JLS3 | Q9jls3 rattus norv | 176 | 7 | 0.5 | 141 | 10 | Q9M2Q5 | Q9m2q5 arabidopsis |
| 104 | 8 | 0.6 | 1311 | 5 | Q9VBQ8 | Q9vbq8 drosophila | 177 | 7 | 0.5 | 142 | 6 | Q9TUB9 | Q9tub9 oryctolagus |
| 105 | 8 | 0.6 | 1311 | 5 | Q9GQNS | Q9gqns drosophila | 178 | 7 | 0.5 | 143 | 2 | Q9RUG3 | Q9rug3 streptococc |
| 106 | 8 | 0.6 | 1312 | 4 | Q9NR59 | Q9nr59 homo sapien | 179 | 7 | 0.5 | 143 | 5 | Q18533 | Q18533 caenorhabdi |
| 107 | 8 | 0.6 | 1331 | 11 | Q9EPQ2 | Q9epq2 mus musculus | 180 | 7 | 0.5 | 144 | 2 | Q9XOW8 | Q9xow8 thermotoga |
| 108 | 8 | 0.6 | 1343 | 4 | Q9H7N4 | Q9h7n4 homo sapien | 181 | 7 | 0.5 | 145 | 4 | Q9NTS3 | Q9nts3 homo sapien |
| 109 | 8 | 0.6 | 1398 | 4 | Q9HCF2 | Q9hcf2 homo sapien | 182 | 7 | 0.5 | 146 | 10 | Q9FNH0 | Q9fnh0 arabidopsis |
| 110 | 8 | 0.6 | 1417 | 10 | Q9FJO9 | Q9fjq9 arabidopsis | 183 | 7 | 0.5 | 148 | 4 | Q9UJX8 | Q9ujx8 homo sapien |
| 111 | 8 | 0.6 | 1520 | 4 | O15087 | O15087 homo sapien | 184 | 7 | 0.5 | 149 | 2 | Q9K9X6 | Q9k9x6 bacillus ha |
| 112 | 8 | 0.6 | 1634 | 5 | Q9XW25 | Q9xw25 caenorhabdi | 185 | 7 | 0.5 | 149 | 10 | Q9FJ67 | Q9fj67 arabidopsis |
| 113 | 8 | 0.6 | 1639 | 4 | Q9H473 | Q9h473 homo sapien | 186 | 7 | 0.5 | 151 | 10 | Q9SDI1 | Q9sdi1 oryza sativ |
| 114 | 8 | 0.6 | 1736 | 4 | O75111 | O75111 homo sapien | 187 | 7 | 0.5 | 154 | 4 | Q9NRN6 | Q9nrn6 homo sapien |
| 115 | 8 | 0.6 | 1781 | 4 | Q9UKX0 | Q9ukx0 homo sapien | 188 | 7 | 0.5 | 154 | 5 | O44604 | O44604 caenorhabdi |
| 116 | 8 | 0.6 | 1857 | 4 | O95153 | O95153 homo sapien | 189 | 7 | 0.5 | 155 | 10 | O65426 | O65426 arabidopsis |
| 117 | 8 | 0.6 | 1871 | 10 | Q9SRD5 | Q9srd5 arabidopsis | 190 | 7 | 0.5 | 156 | 10 | Q9M5F0 | Q9m5f0 medicago sa |
| 118 | 8 | 0.6 | 2073 | 4 | Q9UKW3 | Q9ukw3 homo sapien | 191 | 7 | 0.5 | 156 | 11 | Q9JJF0 | Q9jjf0 mus musculu |
| 119 | 8 | 0.6 | 2073 | 4 | Q9UKW2 | Q9ukw2 homo sapien | 192 | 7 | 0.5 | 157 | 2 | Q9ZLD8 | Q9zld8 helicobacte |
| 120 | 8 | 0.6 | 2248 | 5 | Q9VY17 | Q9vy17 drosophila | 193 | 7 | 0.5 | 159 | 5 | Q9VWNO | Q9vwn0 drosophila |
| 121 | 8 | 0.6 | 3201 | 5 | Q9W0U2 | Q9w0u2 drosophila | 194 | 7 | 0.5 | 159 | 11 | Q9ERD3 | Q9erd3 mus musculu |
| 122 | 8 | 0.6 | 4868 | 13 | Q91319 | Q91319 rana catesb | 195 | 7 | 0.5 | 159 | 14 | Q9PX69 | Q9px69 human herpe |
| 123 | 7 | 0.5 | 35 | 10 | O22245 | O22245 arabidopsis | 196 | 7 | 0.5 | 160 | 4 | Q9POT1 | Q9pot1 homo sapien |
| 124 | 7 | 0.5 | 41 | 2 | O50922 | O50922 borrelia bu | 197 | 7 | 0.5 | 160 | 5 | Q9NLE2 | Q9nle2 leishmania |
| 125 | 7 | 0.5 | 43 | 10 | Q9SQC1 | Q9sqc1 zea mays (m | 198 | 7 | 0.5 | 161 | 1 | O26658 | O26658 methanobact |
| 126 | 7 | 0.5 | 43 | 10 | Q9S704 | Q9s704 zea mays (m | 199 | 7 | 0.5 | 162 | 6 | Q9N1M2 | Q9n1m2 pan troglod |
| 127 | 7 | 0.5 | 48 | 5 | Q9VWS6 | Q9vws6 drosophila | 200 | 7 | 0.5 | 162 | 6 | Q9WYQ2 | Q9wyq2 hylobates l |
| 128 | 7 | 0.5 | 53 | 4 | Q9PLJ2 | Q9plj2 homo sapien | 201 | 7 | 0.5 | 162 | 10 | Q9LWZ3 | Q9lwz3 arabidopsis |
| 129 | 7 | 0.5 | 61 | 3 | Q91627 | Q91627 pneumocysti | 202 | 7 | 0.5 | 166 | 14 | Q9IBS4 | Q9ibs4 turkey herp |
| 130 | 7 | 0.5 | 72 | 4 | O15203 | O15203 homo sapien | 203 | 7 | 0.5 | 168 | 4 | Q9H7G1 | Q9h7g1 homo sapien |
| 131 | 7 | 0.5 | 72 | 6 | O18889 | O18889 ateles belz | 204 | 7 | 0.5 | 170 | 3 | O13569 | O13569 saccharomyc |
| 132 | 7 | 0.5 | 75 | 10 | Q9STQ5 | Q9stq5 arabidopsis | 205 | 7 | 0.5 | 171 | 2 | O85021 | O85021 photobacter |
| 133 | 7 | 0.5 | 81 | 10 | Q9L2K7 | Q9l2k7 arabidopsis | 206 | 7 | 0.5 | 172 | 2 | Q9WYI0 | Q9wyi0 thermotoga |
| 134 | 7 | 0.5 | 87 | 11 | Q63452 | Q63452 rattus norv | 207 | 7 | 0.5 | 172 | 4 | Q9P054 | Q9p054 homo sapien |
| 135 | 7 | 0.5 | 88 | 10 | Q9A412 | Q9a412 brassica ra | 208 | 7 | 0.5 | 172 | 4 | Q9NWD7 | Q9nwd7 homo sapien |
| 136 | 7 | 0.5 | 91 | 2 | P94908 | P94908 microbacter | 209 | 7 | 0.5 | 173 | 2 | O68901 | O68901 salmonella |
| 137 | 7 | 0.5 | 94 | 5 | Q9NNK8 | Q9nnk8 leishmania | 210 | 7 | 0.5 | 173 | 2 | P75600 | P75600 mycoplasma |
| 138 | 7 | 0.5 | 95 | 6 | Q93383 | Q93383 sus scrofa | 211 | 7 | 0.5 | 173 | 5 | Q27189 | Q27189 stylonychia |
| 139 | 7 | 0.5 | 95 | 10 | Q9SCS0 | Q9scs0 arabidopsis | 212 | 7 | 0.5 | 173 | 8 | Q9XKN3 | Q9xkn3 anthophora |
| 140 | 7 | 0.5 | 97 | 11 | Q9WTY6 | Q9wtv6 rattus norv | 213 | 7 | 0.5 | 173 | 10 | O64930 | O64930 pseudotsuga |
| 141 | 7 | 0.5 | 97 | 14 | O64890 | O64890 mouse adeno | 214 | 7 | 0.5 | 175 | 10 | Q9ZT86 | Q9zt86 brassica na |
| 142 | 7 | 0.5 | 98 | 4 | Q9NNW1 | Q9nnw1 homo sapien | 215 | 7 | 0.5 | 177 | 1 | Q9FYZ0 | Q9fyzo aeropyrum p |
| 143 | 7 | 0.5 | 98 | 6 | Q9N106 | Q9n106 macaca mula | 216 | 7 | 0.5 | 177 | 3 | Q9P542 | Q9p542 neurospora |
| 144 | 7 | 0.5 | 103 | 4 | Q9NNZ6 | Q9nnz6 homo sapien | 217 | 7 | 0.5 | 178 | 2 | Q9KFN6 | Q9kfn6 bacillus ha |
| 145 | 7 | 0.5 | 107 | 4 | O15254 | O15254 homo sapien | 218 | 7 | 0.5 | 178 | 10 | O49595 | O49595 arabidopsis |
| 146 | 7 | 0.5 | 107 | 10 | O9LQF6 | O9lqf6 arabidopsis | 219 | 7 | 0.5 | 178 | 10 | O42461 | O42461 canavalia g |
| 147 | 7 | 0.5 | 108 | 2 | Q9R465 | Q9r465 agrobacteri | 220 | 7 | 0.5 | 178 | 10 | O42492 | O42492 canavalia g |
| 148 | 7 | 0.5 | 108 | 4 | O15202 | O15202 homo sapien | 221 | 7 | 0.5 | 179 | 5 | Q9VC01 | Q9vc01 drosophila |
| 149 | 7 | 0.5 | 109 | 1 | Q9REI3 | Q9yel3 aeropyrum p | 222 | 7 | 0.5 | 181 | 5 | O02040 | O02040 caenorhabdi |
| 150 | 7 | 0.5 | 109 | 4 | O15249 | O15249 homo sapien | 223 | 7 | 0.5 | 181 | 6 | Q9TRZ5 | Q9trz5 oryctolagus |
| 151 | 7 | 0.5 | 109 | 4 | O15204 | O15204 homo sapien | 224 | 7 | 0.5 | 183 | 2 | Q9K6F2 | Q9k6f2 bacillus ha |
| 152 | 7 | 0.5 | 109 | 4 | O15200 | O15200 homo sapien | 225 | 7 | 0.5 | 184 | 10 | O22801 | O22801 arabidopsis |
| 153 | 7 | 0.5 | 109 | 6 | O29286 | O29286 sus scrofa | 226 | 7 | 0.5 | 190 | 4 | Q9H2P1 | Q9h2p1 homo sapien |
| 154 | 7 | 0.5 | 110 | 10 | Q9S7Q8 | Q9s7q8 oryza sativ | 227 | 7 | 0.5 | 191 | 2 | Q9ZEB4 | Q9zeb4 rickettsia |
| 155 | 7 | 0.5 | 111 | 5 | Q9U553 | Q9u553 cryptospori | 228 | 7 | 0.5 | 191 | 2 | Q9UEB4 | Q9uel2 homo sapien |
| 156 | 7 | 0.5 | 112 | 10 | O40726 | O40726 oryza sativ | 229 | 7 | 0.5 | 192 | 5 | Q9GUA2 | Q9gua2 riftia pach |
| 157 | 7 | 0.5 | 112 | 10 | Q9FYR1 | Q9fyr1 arabidopsis | 230 | 7 | 0.5 | 193 | 2 | P94221 | P94221 borrelia bu |
| 158 | 7 | 0.5 | 118 | 10 | O65442 | O65442 arabidopsis | 231 | 7 | 0.5 | 193 | 2 | P94227 | P94227 borrelia bu |
| 159 | 7 | 0.5 | 118 | 10 | Q9FP15 | Q9fp15 oryza sativ | 232 | 7 | 0.5 | 194 | 2 | Q9HVC3 | Q9hvc3 pseudomonas |
| 160 | 7 | 0.5 | 118 | 14 | Q9DUD4 | Q9dud4 marek's dis | 233 | 7 | 0.5 | 196 | 2 | Q9KQV4 | Q9kqv4 vibrio chol |
| 161 | 7 | 0.5 | 121 | 4 | Q9R306 | Q9ry3u6 homo sapien | 234 | 7 | 0.5 | 196 | 14 | O55275 | O55275 fowl adenov |
| 162 | 7 | 0.5 | 125 | 4 | Q9H014 | Q9h014 homo sapien | 235 | 7 | 0.5 | 197 | 10 | Q39819 | Q39819 glycine max |
| 163 | 7 | 0.5 | 129 | 10 | Q9LFR4 | Q9lfr4 arabidopsis | 236 | 7 | 0.5 | 197 | 11 | Q9ERC3 | Q9erc3 mus musculu |
| 164 | 7 | 0.5 | 130 | 2 | O50411 | O50411 mycobacteri | 237 | 7 | 0.5 | 198 | 14 | Q9Q829 | Q9q829 shope fibro |
| 165 | 7 | 0.5 | 131 | 2 | O05787 | O05787 mycobacteri | 238 | 7 | 0.5 | 198 | 14 | Q9IGT8 | Q9igt8 porcine ade |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2001, 10:04:49 : Search time 40.2 seconds
(without alignments)
4578.017 Million cell updates/sec

Title: US-09-512-581-2
Perfect score: 1391
Sequence: 1 MAHSKTRNDGKITYPGVK.....QKGRGRPSKTPSPSQPKKNV 1391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_16.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_unclassified.*
13: sp_vertebrate.*
14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 1015 | 73.0 | 1416 | 4 | Q9v215 homo sapien |
| 2 | 1015 | 73.0 | 1447 | 4 | Q9nt15 homo sapien |
| 3 | 353 | 25.4 | 1391 | 4 | Q9t451 homo sapien |
| 4 | 125 | 9.0 | 125 | 4 | Q9h5N8 homo sapien |
| 5 | 19 | 1.4 | 851 | 4 | Q9v4d4 homo sapien |
| 6 | 9 | 0.6 | 419 | 4 | Q93029 homo sapien |
| 7 | 9 | 0.6 | 554 | 10 | Q9SG14 |
| 8 | 9 | 0.6 | 837 | 3 | Q9P607 |
| 9 | 8 | 0.6 | 59 | 11 | Q9R1N5 |
| 10 | 8 | 0.6 | 66 | 11 | Q9WVP2 |
| 11 | 8 | 0.6 | 81 | 10 | Q9M3G5 |
| 12 | 8 | 0.6 | 99 | 5 | Q23793 |
| 13 | 8 | 0.6 | 99 | 5 | Q23794 |
| 14 | 8 | 0.6 | 131 | 10 | Q96475 |
| 15 | 8 | 0.6 | 147 | 14 | Q84290 |
| 16 | 8 | 0.6 | 149 | 10 | Q9LV56 |
| 17 | 8 | 0.6 | 197 | 5 | Q23872 |
| 18 | 8 | 0.6 | 204 | 10 | Q9SB54 |
| 19 | 8 | 0.6 | 231 | 3 | Q9P529 |

| | | | | | |
|----|---|-----|------|----|--------|
| 20 | 8 | 0.6 | 235 | 10 | Q9SXG4 |
| 21 | 8 | 0.6 | 251 | 4 | Q95472 |
| 22 | 8 | 0.6 | 294 | 10 | Q22924 |
| 23 | 8 | 0.6 | 299 | 10 | Q9M4F1 |
| 24 | 8 | 0.6 | 310 | 10 | Q39175 |
| 25 | 8 | 0.6 | 313 | 10 | Q39089 |
| 26 | 8 | 0.6 | 316 | 2 | Q50887 |
| 27 | 8 | 0.6 | 316 | 10 | Q9LI19 |
| 28 | 8 | 0.6 | 325 | 5 | Q44016 |
| 29 | 8 | 0.6 | 329 | 10 | Q9FM77 |
| 30 | 8 | 0.6 | 331 | 5 | Q9VD43 |
| 31 | 8 | 0.6 | 341 | 1 | Q9YFQ6 |
| 32 | 8 | 0.6 | 351 | 5 | Q9G263 |
| 33 | 8 | 0.6 | 368 | 3 | Q9P837 |
| 34 | 8 | 0.6 | 379 | 10 | Q9XE25 |
| 35 | 8 | 0.6 | 388 | 6 | Q9GL18 |
| 36 | 8 | 0.6 | 389 | 6 | Q9GK25 |
| 37 | 8 | 0.6 | 395 | 10 | Q9LTR6 |
| 38 | 8 | 0.6 | 425 | 4 | Q9NQD6 |
| 39 | 8 | 0.6 | 442 | 2 | Q34631 |
| 40 | 8 | 0.6 | 450 | 11 | Q9WVR7 |
| 41 | 8 | 0.6 | 462 | 11 | Q9R1B7 |
| 42 | 8 | 0.6 | 466 | 3 | Q9HEL6 |
| 43 | 8 | 0.6 | 468 | 1 | Q57853 |
| 44 | 8 | 0.6 | 470 | 11 | Q9QYK4 |
| 45 | 8 | 0.6 | 475 | 4 | Q43818 |
| 46 | 8 | 0.6 | 501 | 10 | Q39482 |
| 47 | 8 | 0.6 | 505 | 11 | Q37770 |
| 48 | 8 | 0.6 | 507 | 10 | Q9SL49 |
| 49 | 8 | 0.6 | 510 | 4 | Q75239 |
| 50 | 8 | 0.6 | 519 | 4 | Q95676 |
| 51 | 8 | 0.6 | 521 | 11 | Q08795 |
| 52 | 8 | 0.6 | 522 | 5 | Q61090 |
| 53 | 8 | 0.6 | 531 | 4 | Q9H166 |
| 54 | 8 | 0.6 | 535 | 11 | Q9JY99 |
| 55 | 8 | 0.6 | 545 | 4 | Q9H117 |
| 56 | 8 | 0.6 | 552 | 10 | Q49397 |
| 57 | 8 | 0.6 | 562 | 3 | Q9P557 |
| 58 | 8 | 0.6 | 564 | 10 | Q43673 |
| 59 | 8 | 0.6 | 566 | 2 | Q9WZU1 |
| 60 | 8 | 0.6 | 568 | 2 | Q9Z851 |
| 61 | 8 | 0.6 | 568 | 2 | Q9JSE4 |
| 62 | 8 | 0.6 | 583 | 11 | P70616 |
| 63 | 8 | 0.6 | 587 | 10 | Q9LWV1 |
| 64 | 8 | 0.6 | 587 | 11 | P70615 |
| 65 | 8 | 0.6 | 588 | 11 | Q61791 |
| 66 | 8 | 0.6 | 602 | 14 | Q56852 |
| 67 | 8 | 0.6 | 605 | 4 | Q9NZ53 |
| 68 | 8 | 0.6 | 612 | 13 | Q919E7 |
| 69 | 8 | 0.6 | 644 | 4 | Q9H3E6 |
| 70 | 8 | 0.6 | 646 | 10 | Q64877 |
| 71 | 8 | 0.6 | 655 | 2 | Q9KQR4 |
| 72 | 8 | 0.6 | 662 | 5 | Q9VFD5 |
| 73 | 8 | 0.6 | 665 | 4 | Q76034 |
| 74 | 8 | 0.6 | 666 | 10 | Q9FK63 |
| 75 | 8 | 0.6 | 671 | 3 | Q12242 |
| 76 | 8 | 0.6 | 711 | 4 | Q9U2H3 |
| 77 | 8 | 0.6 | 720 | 5 | Q9U2H3 |
| 78 | 8 | 0.6 | 730 | 10 | Q82332 |
| 79 | 8 | 0.6 | 765 | 10 | Q9FZA9 |
| 80 | 8 | 0.6 | 787 | 10 | Q9M0L6 |
| 81 | 8 | 0.6 | 791 | 13 | Q9DGL1 |
| 82 | 8 | 0.6 | 802 | 10 | Q38898 |
| 83 | 8 | 0.6 | 840 | 13 | Q918N3 |
| 84 | 8 | 0.6 | 841 | 1 | Q59612 |
| 85 | 8 | 0.6 | 874 | 11 | Q88453 |
| 86 | 8 | 0.6 | 938 | 4 | Q13187 |
| 87 | 8 | 0.6 | 954 | 4 | Q15326 |
| 88 | 8 | 0.6 | 973 | 4 | Q9H0Z1 |
| 89 | 8 | 0.6 | 1037 | 10 | Q9FH58 |
| 90 | 8 | 0.6 | 1047 | 3 | Q9P6E3 |
| 91 | 8 | 0.6 | 1062 | 4 | Q9NSW2 |
| 92 | 8 | 0.6 | | | |

Q9sxg4 oryza sativ
Q95472 homo sapien
Q22924 arabidopsis
Q9m4f1 brassica na
Q39175 arabidopsis
Q39089 arabidopsis
Q50887 myxococcus
Q9li19 oryza sativ
Q44016 dictyosteli
Q9fm77 arabidopsis
Q9vd43 drosophila
Q9yfg6 aeropyrum p
Q96263 plasmodium
Q9p837 candida alb
Q9xe25 oryza sativ
Q9gl18 diceros bic
Q9gkz5 tupala tana
Q9ltr6 arabidopsis
Q9ncd6 homo sapien
Q34631 bacillus su
Q9wvr7 rattus norv
Q9rib7 rattus norv
Q9hel6 neurospora
Q57853 pyrococcus
Q9yk4 mus musculus
Q43818 homo sapien
Q39482 calocedrus
P97770 mus musculus
Q9sl49 arabidopsis
Q9va02 drosophila
Q75239 homo sapien
Q95676 homo sapien
Q08795 mus musculus
Q61090 plasmodium
Q9h166 homo sapien
Q9jy99 mus musculus
Q9h117 homo sapien
Q49397 arabidopsis
Q9p557 neurospora
Q43673 vicia faba
Q9wzui thermotoga
Q9z851 chlamydia p
Q9jse4 chlamydia p
P70616 rattus norv
Q9lwl1 oryza sativ
P70615 rattus norv
Q61791 mus musculus
Q66852 fowl adenov
Q9nz53 homo sapien
Q919e7 brachydanio
Q9h3e6 homo sapien
Q64877 arabidopsis
Q9kqr4 vibrio chol
Q9vfd5 drosophila
Q76034 homo sapien
Q9fk63 arabidopsis
Q12242 saccharomyc
Q9h116 homo sapien
Q9u2h3 caenorhabdi
Q82332 arabidopsis
Q9fza9 arabidopsis
Q9m0l6 arabidopsis
Q9dgl1 fuqu rubrip
Q38898 arabidopsis
Q918n3 rana pipien
Q59612 pyrococcus
Q88453 rattus norv
Q13187 homo sapien
Q15326 homo sapien
Q9h0z1 homo sapien
Q9fh58 arabidopsis
Q9p6e3 neurospora
Q9nsw2 homo sapien

med sep 20 11:11

12754 CCACAAAACAAGTCAGGTGAAGACTCTACGGATAATTTTACAAAGGAAAA 121705
1058AspAlaLysMetAsnGluLysLeuTyThr 1067
121704 GGCAATTTTCCAATTGGAACACAGATTAATCAAGAGCAAGAAGACCGAAC 121655
1067 rValCys.....AspValAlaMetAsnIlelleMetSerL 1079
121654 ACTTACCAGCCAGTATCTGGAGATACATCTTAACAAAACACAGAGTGCA 121605
1079 ysSerThrThrTyThrSer.....LeuGluSerProLysAspProVal 1092
121604 CTCTGCAGTTGATTCGGATCAGCATTTGTTATCCGTTAAGAGTAGAGCT 121555
1093 LeuProAlaArgPhePheThrGlnPro...AspLysAsnPheSerAsnTh 1108
121554 GGGCTAGCCAG.....GACAAACCATCTGAGAGTGCTATTTCATAGTAC 121511
1108 rLysAsnTyThrLeuProGluMetLysSerPhePheThrProGlyLys. 1124
121510 TTCACAGACTCGAAAAAGTAAGAACACACAGACATTAACAACCTTCCAAC 121461
1125ProLysThrThrAsnValLeuGlyAlaValAsnLysPro 1137
121460 GAACAATATTCACAGAGTCCACAGAGATGTCGGGACTTATCAGCAGCCT 121411
1138 LeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArg.....1151
121410 TCNAACCTCAGAAAAAGTTAAACCCCATGAAGAGAAATCTTAACACCTAAA 121361
1152MetGluThrValSerAsnAlaSerSerSerSerSerAsnPros 1165
121360 AAGAGTGAACAAAAAGGGATCTCTCCGCTGATGGGGGACTGAGGCTG 121311
1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGluMet.....1179
121310 AACCTCCAAAGAGAGTG....TCGAGGAACCATCTGCTGAGGTTTGGAA 121264
1180AspHisSerGluAsnGluAspTyThrMetSe 1190
121263 ATAGAAGACCATGGAAGCCACACTAAGCAAGAGATGATCACAGATTAA 121214
1190 rSerProLeuProGlyLysLysSerAspLysArgaspSerAspLeuv 1207
121213 ATCTAGGGGGCGAGCTAGAAAACCATCATCATCATCATCATCATCATCAT 121164
1207 alArgSerGluLeuGluLysProArgGlyArgLysLysThrProValThr 1223
121163 AGGAAGCTATTACAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121114
1224 uGlnGlnGluLysLeuGlyMetAspLeuThrLysLeuValGlnGl 1240
121113 GAACCTTCTGAA.....ACCAGCAAAATGTCGA 121085
1240 uGlnGlnProLysGlySerGlnArgSerArgLysArgGlyHisThrAlas 1257
121084 ACAGAAA.....GAATCCAGGCCAT.....121063
1257 erGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAsp 1273
121062 ..AAATCTGATGAAGCTCTTAACCC.....121039
1274 IleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysAr 1290
121038 ATCCTTAGAAGAACACACAGAG.....CCAGAGTTGAGAAAG 121001
1290 q.....GlyArgProPolypSP 1296
121000 ANTGAGTCAAATGCAGAACAGGTGATCATCATCATCATCATCATCATCATCAT 120951
121000 ANTGAGTCAAATGCAGAACAGGTGATCATCATCATCATCATCATCATCATCAT 120951

THIS PAGE BLANK (USPTO)


```

334  ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350
      |||||: : : : : |||:
124094  ....TGCCTTTGTGGAACAAAAAC.....AAGAA 124069

350  pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleA 367
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
124068  CTTGATG...TTTCTACTGATGCG..... 124047

367  rghisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeu 383
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
124046  ....AATCAATCAACATTGCATCGAAAGTTCCGCTC 124014

384  LeuValAsnAspHisLeuLeuAsnPheValArgGluArgThr..... 397
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
124013  TCAACGAACATTCAGAGAATATCCAACTGAGGAGGTACACACTTTAGAT 123964

398  ....LeuAspLysArg.....TrpArgVal...ArgL 406
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123963  AATACTAACAAGATAAAACGGTATCCCGTCATCGGAAGTTTAAAGGA 123914

406  ysGluAlaMetMetGlyLeuAla..... 413
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123913  AACAAAACTGATGATCATCTCCAAATCCGTCACCAATACGTATCCAAACG 123864

414  ...GlnIleTyrLysLysTyrAlaLeuGlnSerAla..... 424
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123863  AGGAACCTTCAAGAAGAAGTAAAGCTGATGAGTTGCCTACTGTTAGTAA 123814

425  ....AlaGlyLysAspAlaAlaLysGlnIleAlaTyrPileLysAspL 439
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123813  ATGTTCAGACATTCAAAGACAGACAAACAGATGATCTATTAAACCAAT 123764

439  ysLeuLeuHisIleTyrTyrGlnAsnSer.....Ile 449
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123763  CATCGAAACATCGCTGGAAGAGACCTTAGAAGAGGCACATCCTGATG 123714

450  AspAspArgLeuLeuValGluArgIlePhe..... 459
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123713  AGTTGCTTCTTCTAGTGAAGAGGATCAAGTAATTTGCGAAACACCATCT 123664

460  ....AlaGlnTyrMetValProHisAsnLeuGluT 470
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123663  GCTGAACATAGCCCAACCATCGAAACTCTCTGTGACGAACCTATCAAC 123614

470  hrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123613  AGCAGCAACTGATATAGTCCTGATT..... 123588

487  AsnAlaValLysAlaLeuAsn.....Glu 494
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123587  ...TAAGTCAAAATTGATCACTAAATTTTGAANAATCACACGCCGAAGAA 123541

494  uMetTyrLysCysGlnAsnLeuArg..... 503
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123540  CCGCATGAGAACAACAACTGATGAGACTACCCAGGAAGTCAATCGAAAC 123491

504  ..His..GlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrA 519
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123490  ATCACACGCCGGAAGAAGATCAATAGAA.....GCCAGTCTTAAGAGG 123447

519  spAlaSerValLysAlaIlePheSerLysValMetValIle..... 532
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123446  ATTCACTGTGTAGTAATAATGGATAATGCAGTTGTGGAAACTAATCCTCAA 123397

533  ThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPh 549
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123396  ACAGATAAGCTAATTCCTCCGGATGAATACAAATTGCGGGCCG..... 123352

549  eThrGlnValLeuGluAspGluLysIleArgLysGln.....L 563
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123351  .GTAGATATATGATCGATGAGTCTCTTTAAGAAATTTCTAGTTACGAT 123303

```

```

563  euGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCys 579
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123302  TTTCAGATCGAACACTCAGAAAGTCGCGCAGCACTTCTTTAGACTTAGAT 123253

580  ValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnPr 596
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123252  CTTTCGAGATAAGTAAGAAAGGAATATGGAAGCTTTGCCGAAT.. 123205

596  oPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisI 613
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123204  ....GACAACGCTTGTGTTAAGCACCA 123183

613  le.....AspThrGluSerIleSerAlaLeuIleLysGlnValAsnLys 627
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123182  TAGCGTGCATCTTCTAATGTAAGCTCATTAAGAACCAACGAATTTA 123133

628  Ser.....IleAspGlyThrAlaAspAspGluAspGluGlyVa 640
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123132  TCTCCCAAGACTCGTCTAGCTGGACATCAATGTCAAGATGAGGCCGC 123083

640  lProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysValLeus 657
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123082  TCAAGTTGATGAACAG.....A 123066

657  erPheThrHisProIleSerPheHisSerAla...GluThrPheGluSer 672
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123065  ATGAAAACGATCAGCTCGAAGAACACGTCGCAATGGAGGTATCATCAAT 123016

673  LeuLeuAlaCysLeuLysMetAspAspGluLys.....ValAlaGl 686
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
123015  GAAATCGCTGCATAAAGAGATGATGAGAAATTTGAAGACATAGATGA 122966

686  uAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAsp 703
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122965  GCATGCTTGGCGGAGTTAAACAA.....AGTTCACCTGGAAGAAAT.. 122923

703  heProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLys 719
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122922  ....GCATTAATTCATCAATCTCTA 122902

720  LysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaI 736
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122901  ACGGACCAGAATAAGAGGTTTCTATTATGGAAGAATCCGAAGCTGTTGT 122852

736  ePheSerSer...LysGluThrGlnPheAlaGlnIlePheGluProLeuH 752
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122851  TTTAGAGGAACCTAAGGACAGCAACCAAAATCAGGATTTCAAAGCAGAG 122802

752  isLysSerLeuAspProSerAsnLeu.....GluHisLeuIleThrPro 766
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122801  TGATTCGCTTAGACGCTCCATCATTTGTTGAGAAAAGATACACTAGATGA 122752

767  LeuValThrIleGly.....HisIleAlaLeuLe 776
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122751  AGTACATCAGTTGGAGAGGATAAGATCAATCCACAGAANTTTCTATGCT 122702

776  uAlaProAspGlnPheAlaIaProTyrLysSerTrpValAlaIaThrPheI 793
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122701  TAAAGAACAACTCTAATGCGGACCAA..... 122677

793  leValLysAspLeuLeuMetAsnAspArg.....Leu 803
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122676  ..ATAAAGATGTTAGCTCGATGGCAAAAAAAGAATTCACAGAATGTATA 122629

804  ProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPr 820
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122628  CCAAAAAAGGATTCT.....GACGAAAAAGCAANTGA 122597

820  oGluThrMetValLysIleGlnAla..... 828
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||
122596  CGATAAAGTTATTCAATTAGATGCTCGAGCTTAGATGAAAGTTACACAG 122547

829  .....IleLysMetMetValArg 834

```


[illegible]


```

alignment_block:
US-09-512-581-2 x SCRIF1
Align seg 1/1 to: SCRIF1 from: 1 to: 6541

3 HisSerLysThrArgThrAsnAspGlyLysIleThrTyr..... 15
: : : : : : : : : : : : : : : : : : : : : : : : :
1000 TATAAGGAACAAGGTAATCCTCGCGCTTAGAGTATTGGTGAGTGG 1049

16 .....ProProGlyValLysGluIleSerAspL 25
: : : : : : : : : : : : : : : : : : : : : : : : :
1050 TGAGATCCATGCTTTAGTTGATAAAGAAAGTGTTCGGAGTTTAAAGAAA 1099

25 ysile.....SerLysGluGluMet 31
: : : : : : : : : : : : : : : : : : : : : : : : :
1100 TTATTGAGGCGGATTTGGCATCTTGGACACAGAATCTGAAGATTACGTC 1149

32 ValArgArgLeuLysMetVal.....Vally 40
: : : : : : : : : : : : : : : : : : : : : : : : :
1150 GCGAGGAGGTTTGGAGGTATACGAACCTTTTAATAATATTATACCCATTCT 1199

40 sThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrL 57
: : : : : : : : : : : : : : : : : : : : : : : : :
1200 ACCCACTAAATATTAAGTAGGTGATCAGAAATTCAAATATTCTAATG 1249

57 euAsnLeuAla.....LeuHisLeuAla 64
: : : : : : : : : : : : : : : : : : : : : : : : :
1250 TGAATATAGAAAAGTATCATTTGAAATATGCATACCGCATCTACAAATAGCA 1299

65 SerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 81
: : : : : : : : : : : : : : : : : : : : : : : : :
1300 CAGGACACATTTAAGTAGTAGTGAAGAAAAAATCCATTGTGAATC... 1347

81 acYscYsLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrT 98
: : : : : : : : : : : : : : : : : : : : : : : : :
1348 .....AGACTATATGTTCAATAGTTCGATTCT 1375

98 hrSer..... 99
: : : : : : : : : : : : : : : : : : : : : : : : :
1376 TCAGTGCTATTATGCTAATTTCAAAATCGTTAAGTGGCTAACGAAAGG 1425

100 ProAsp.....LysLeuLysAspIlePhe..... 107
: : : : : : : : : : : : : : : : : : : : : : : : :
1426 CCAGATTGGTTTAAGTAAAGTAAAGTATTATTACAGGTGGACACCGGTGC 1475

108 .....MetPheIleThrArgGlnLeuLysG 116
: : : : : : : : : : : : : : : : : : : : : : : : :
1476 GCTAGAATGAAATTCACAAAGATATATCATACAGCACAGTCTCTT 1525

116 lYLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeu 132
: : : : : : : : : : : : : : : : : : : : : : : : :
1526 TCTTGAGAGACCAAAA.....TTGGTACTTTTTTCTTGAGCAAC 1566

133 GluAsnIle...AlaTrpValLysSerTyrAsnIleCysPheGluLeuGl 148
: : : : : : : : : : : : : : : : : : : : : : : : :
1567 GAAGAAATTTAAACCTATCATTTAGCACTTTCACTGAAATAAATGGAATCAA 1616

148 uAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValI 165
: : : : : : : : : : : : : : : : : : : : : : : : :
1617 TAGTCACACACCTAATATATGAAAAATG.....CTAC 1648

165 leAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMet 181
: : : : : : : : : : : : : : : : : : : : : : : : :
1649 TTAAGGGGATTTTGGACCAAGTATCCAAAATTAATGATTGAGACGGTA 1698

182 SerSer.....IleIleCysGluGlyAs 189
: : : : : : : : : : : : : : : : : : : : : : : : :
1699 ACTTCTTGTTTACCAGGTGAGGTACTACTAGGATTTATTCGGAGATGA 1748

189 pThrValSerGlnLeuLeuAspThrValLeuValAsnLeuValProA 206
: : : : : : : : : : : : : : : : : : : : : : : : :
1749 AATTTACTCCATGAATTTCTCATACTTCAATAGTTGTTTACTACG.... 1794

```

```

206 laHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeu 222
1795 .....GAACTACTA 1803

223 LysArgThrAlaGlnAlaIleGluProTyrIleThrPhePheAsnGl 239
: : : : : : : : : : : : : : : : : : : : : : : : :
1804 AAAAAAGCTTAGATTTTGTGTAAGCAATGAAGGATTAC.....CA 1847

239 nValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValp 256
: : : : : : : : : : : : : : : : : : : : : : : : :
1848 GTGTATCATGCTATCTCCAGTATCGAAACAATCCCGGAAAAATTTTAT 1897

256 heAspLeuIleLeuGluLeuTyr.....AsnIleAsp..... 266
: : : : : : : : : : : : : : : : : : : : : : : : :
1898 CTAACACTACCGTTAAATTCATATGACAGCGCTAATTAGATAAAGTGACA 1947

267 ....SerHisLeuLeuSerValLeuProGlnLeuGluPheLysLeuLy 282
: : : : : : : : : : : : : : : : : : : : : : : : :
1948 ATGTGTCATCTCTTACACAGCAGATAAAA.....AATATATTGTGGT 1991

282 sSerAsnAspAsnGluArgLeuGlnVal...ValLysLeuLeuAlaL 298
: : : : : : : : : : : : : : : : : : : : : : : : :
1992 CAAAATGACATAAATAAGCAATGGATTTGTGGCTTTCATGACCGGAT 2041

298 ysmetPhe.....GlyAlaLysAspSerGluLeuAlaSerGlnAsnLys 312
: : : : : : : : : : : : : : : : : : : : : : : : :
2042 TACTATACGATAGCGAAAGAGAGATTTATGATTGACATCAGAAATCTAAC 2091

313 ProLeuTrp.....GlnCysTyrLeuGlyArgPheAsnAs 324
: : : : : : : : : : : : : : : : : : : : : : : : :
2092 AAAGTTTGGTTTGATTTAAATAACCTGTCTTTATCAATAAT..... 2133

324 pIleHisValProIleArgLeuGluCysValLys.....PheAlas 338
: : : : : : : : : : : : : : : : : : : : : : : : :
2134 ....CACCCCAAGACTCGCTTAATGTCCATCAAGTTTGGAGAAATTA 2179

338 erHisCysLeu.....MetAsnHisProAspLeuAlaLys 349
: : : : : : : : : : : : : : : : : : : : : : : : :
2180 CCTACTGCATATCCACGAAATATCACAGAAAAATCAGGAAGGAAACAA 2229

350 AspLeuThrGluTyrLeuLys.....Va 357
: : : : : : : : : : : : : : : : : : : : : : : : :
2230 AGTTTACTGTCTACTTTTACGAACCCCTTTTCAAATGACGCTTCTTATGT 2279

357 lArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerI 374
: : : : : : : : : : : : : : : : : : : : : : : : :
2280 GAATGATCCAACCGCTAGAGAAGGATCATCTACCACCTTTTAGGTGTTG 2329

374 leValThrAla.....AlaLysLysAspIleLeu..... 383
: : : : : : : : : : : : : : : : : : : : : : : : :
2330 TTTACACTGCATTTACAAGTAATAAGAATTTGCAACAGATATGTTTGA 2379

384 LeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLy 400
: : : : : : : : : : : : : : : : : : : : : : : : :
2380 CPTATTTGGGACCACCTAATAACCCCGATTTATGAAGATTATGTTTCAA 2429

400 sArgTrpArgValArg..... 405
: : : : : : : : : : : : : : : : : : : : : : : : :
2430 ATATGATTCGATTCTACTTCCAAAATGTACTCTTCACAGTTTTCATTGT 2479

406 .....LysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
: : : : : : : : : : : : : : : : : : : : : : : : :
2480 TAATTTGGGGAAGAATGACAGATGTTGATTGGAAAGAAATATAAAAG 2529

419 Tyr.....AlaLeuGlnSerAlaAlaGlyLysAs 428
: : : : : : : : : : : : : : : : : : : : : : : : :
2530 CATATTCACCCCAATGAGTGTATAGCCTCAGAAAGGTGTAAGCTTAAAGA 2579

428 pAlaAlaLys.....GlnIleAlaTrpIleLys.....AspL 439
: : : : : : : : : : : : : : : : : : : : : : : : :
2580 CATTTCAAGCTGCCCGCCGACATC.....ATCAAGCGTGAATACGATA 2623

```



```
1179 tAspHisSerGluAsnGluAspThrMetSerSerProLeuProGlyL 1196
1178 TAATGAGGGTGAACAAACAGATTATGCTTGAATGTT..... 1142
1196 yLysSerAspLysArgAspSerAspLeuValArgSerGlu..... 1210
1141 .....GACCAAGCTCTCCAGTAAGGATGAGTGCCA 1109
1211 .....LeuGluLysProArgGlyArgLysLysThrProValThrG 1224
1108 GGAGACGGCATTGAGTCCGAGTGGGACTATCTTCTCAATGTTCAA 1059
1224 uGlnGluLysLeuGlyMetAspAspLeuThrLys..... 1236
1058 GCAGACAGAAAAATCAAAAGTTGATGCTTGGCTAGTGAAGATGAAC 1009
1237 .....LeuValGlnGluLysProLysGlySerGlnArgSerArg 1250
1008 ATGGAACGGTGGCGAGAGAAACATCAAGTGGTGCATTAACAGTAGA 959
1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGlu 1267
958 AACAAAAAATATGACAGGATGGACAGCACACCCATTCAGGCACCTGA 909
1267 uLysArgLeuLysGluAspLeuGluAsn.....G 1278
908 AGAATCTAGGAGAGTGTGTTATGACCGAAGAGGTATCAATGTACGCT 859
1278 luAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPro 1294
858 AAGATAGTGAAACATCGCAACTTAATAAAATTTGAAGGCA..... 818
1295 LysProLeuGlyGlyGly..... 1300
817 ...CCGCTTAAGGAGATAGGATGCAATATTAATGACGATTTGTGACC 771
1301 .....ThrProLysGluGluProThrMetLysThrSerLysLysG 1314
770 TGTGTAAGAAAATGTCAGAGATGAGGCTTTTAAATCAATGGAACAT 721
1314 LysSerLysLysSerGlyProProAlaProGluGluGluGluGlu 1330
720 CGGTACTAAGAGACAGGT.....CTAGAAGACACACC 686
1331 Glu..... 1331
685 GAAGTGTCTGATATTTCAGGTTCACGAGAAATTCGTATACCCATTTTCAA 636
1332 .....ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGln 1346
635 TTCTTTGAAATCAAGGGTCAAGAGACCCAAATAAAGAGAAAAATTGAAA 586
1346 isArgValSerArgAlaGlnGlnArgAlaGluSerProGluSerSer 1362
585 AGAGTTGCAAGAGATGAGTTAATGCCACCATTCACCTCTCTCGAATG 536
1363 AlaileGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1379
535 ACA...GAAAAATCAAAATATCAACGGCGACAGAAATGCTTAGACAC 489
1379 LysThr 1381
488 AAAAAACA 482
```

seq name: gb_p13:SCRI1

seq_documentation_block: 6541 bp DNA 11-AUG-1992
LOCUS SCRI1
DEFINITION S.cerevisiae RIF1 gene.
ACCESSION X66501

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
Source

gene

CDS

X66501.1 GI:4327

RIF1 gene.

Saccharomyces cerevisiae.

Saccharomyces cerevisiae.

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

Hardy,C.F., Sussel,L. and Shore,D.

A RAPI-interacting protein involved in transcriptional silencing

and telomere length regulation

Genes Dev. 6 (5), 801-814 (1992)

2 (bases 1 to 6541)

Shore,D.M.

Direct Submission

Submitted (30-JUL-1992) D.M. Shore, College of Physicians &

Surgeons of, Columbia Univ., Dept. of Microbiology, 701 West 168th

Street, New York, N.Y. 10032, USA

Location/Qualifiers

1. 6541

/organism="Saccharomyces cerevisiae"

/strain="AB320"

/db_xref="taxon:4932"

454..6204

/gene="RIF1"

454..6204

/gene="RIF1"

/codon_start=1

/protein_id="CAA7121.1"

/db_xref="GI:4328"

/translation="MSKDFSDKKHTIDRQDILRRSHQDNYSNGSPWMTNLP

SPQAHMHQSDLSPTPKRRKLASSDCENKQDFDLSAINKNLYPDTGSRMQLS

ASNDNVSPVTSVAFSDRIESSPIIRPGSPKSPGSKGSLNRPVSRVSD

LSYNLQYTHKLHNGNFTSPYKTRVPRALRYVSGEHLNRPVSRVSD

GGLILROESEDYVARFEVATFNIIPIITLKNNVNDQKFNILIVNIESI

PHLQIAOTLLSSSEKPNFVRLVQIVRFSAIMSNFKIVKTKRPLDNVNLKI

RYWTTGALRNENSKIIITAOVSFLRDEKFGTFILSNBEIKPILSTFTEIM

YKELLIRGFLSKYKPLMIETVTSWLPCEVLPRIIIGDEIYSMLILTSIV

KKDFVDEHERIYQCIIMLSPVCTEIPKFLSKPLNSYDNLKRVTHGLTQ

NIYVNDNKIAMDLWSMTGLLYDSGRVYDLTSESNKVFDMNLNLCFIN

MSIKVRIITYCISTIKSQKQEGNKLSSLLRTPTQMTLPTVNDPSAREGI

VVYTAFTSNKLSLDMFLFDHLITPIYEDYVKYDSTHLQNVLTFLHLLIG

VALERKVKHHPMSVIASGVKLKDISLIPQIIRKREYDKIMKVVFAVAT

NLAHDLILTSLKHLPEDRKDOOTHEFSLSLLKVTQNNKQDPIPROFGA

FDLFLRNDSLSYFNFIQISKVIGSQGNWLDLLKRVKARNETSEFLIEK

DKKTEVYANQWVGSTLPPNISREFOSLANIVNKNPNENSIEFLLDLK

LFTLLHVSWSNNNFYFIQSVSKNENKLVNLDLTLKTSLPNGPELFG

NKFMDELYCIHSPNLLNSIPDLNSDLLKLLPRSRASYFAANIKLFC

WLLKGOOLEQNLONFSEIENVLONASDSELEKSEIIRLLHLAMANPTE

CIKNMADHLEFCGNTSEVLFKLSPELLLLKLTYSKPNGLAAVTEK

YILELEKIIIOKEIQIILEKKEPLGVFLNPNMOKHKKASTNMLREL

LSRSAAKFFSMLISILPPNPNYQITDMVNLIDILKSHNRKFKDRTY

WQESGVVHQGDSKEIEIAPDTSKSWYIPCESENKLNLRKVSDQIQVAT

EPSPKIQSSOISAKNDSISLKNATIMNSQOESHANRSIDDETLEEVN

IDQOMKSTQDKNVANHCISCTKSDVDVTELSHESIDTOSSEVAYOP

AVTNRSIKTNDHNVNDSNPLKRPSETPTSENKRSKGTHETVMDVLY

DVICTNIKIANEESLALNSIKVETNCNLSNLTLDQOQITKEDGKGQ

ROENQSNKINSKFTQNTAQYKSKARPNEGNNNDYACNVQSPVREY

GIQIPSGTILLNSSKQTEKSKVDLRSDEDEHGTVAQKHQVGNKIN

PIQGTESREVMTEEGINVRLEDGTCCLNKLKPLKGDKNANDINRND

RDEGLKSMHAKVSKETGLEQPEVADISVLPETIRIPFNSLKMGSQ

LQNELMPPDSPPRMTENTNINAQGLDTPVKITGCKEKKHEIQQAHT

GGGDNEDATREATPSLKVHFFSKSRRLVAGTTPGDLNGLISVEER

FMRLLEYYSNRDNDMN"

BASE COUNT 2340 a 1116 c 1224 g 1861 t

ORIGIN

alignment_scores:

Quality: 226.00

Ratio: 0.292

Percent Similarity: 42.496

Length: 1819

Gaps: 80

Percent Identity: 17.372

| | | |
|------|---|------|
| 3682 |TTGAAGCATCTT | 3671 |
| 506 | ValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerVa | 522 |
| | | |
| 3670 | CCAGAGGAT.....AGAAAGCAACAACTCATTT | 3642 |
| 522 | llyslalallepheSerLysValMetValIleThrArgAsnLeuProAspp | 539 |
| | | |
| 3641 | AGAATCCTTTAGCAGCTCTATATTAAAGTCACGCAAAACAAACAAGATA | 3592 |
| 539 | roGLyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu..... | 553 |
| | | |
| 3591 | CACCACATTTTCGTGACTTTTTTGGTGCTGTACTTCATCGTTGGTTTAT | 3542 |
| 554 |GluAspAspGluLysIleArgLy | 561 |
| | | |
| 3541 | ACATTTTGGACCTGTTTTTAAGAAGNATGATTCATCATTAGTAATTT | 3492 |
| 561 | sGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluG | 578 |
| | | |
| 3491 | CAATATCCAAATTTCCAAAGTTGGTATTTCGCAAGGAACATGACCCCTTG | 3442 |
| 578 | lYcysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThr | 594 |
| | | |
| 3441 | ATCTTTTAAAAGATGTTATCAGGAAGCTCGAAAC.....GAAACA | 3401 |
| 595 | AsnProPheLeuGluMetIleLysPheLeuLeu..... | 605 |
| | | |
| 3400 | TCGGAATTTCTGATTAATAGAAAGTTTCTAGAAATGGATGATAGAAAAAC | 3351 |
| 606 |GluArgIleAlaProValHisI | 613 |
| | | |
| 3350 | AGAAGTATATGCCAAAATTTGGTGGATCAACGTTATTACCACCAATA | 3301 |
| 613 | leAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIle | 629 |
| | | |
| 3300 | TATCTTTCAGAGAATTTTCAGAGTTTGGCGAATATGTGCAACA..... | 3257 |
| 630 | AspGlyThrAlaAspGluAspGluGlyValProThrAspGlnAlaIle | 646 |
| | | |
| 3256 |GTCCGAATGAAATTCAT | 3237 |
| 646 | eArgAlaGlyLeuGluLeuLysValLeuSerPheThrHisProIles | 663 |
| | | |
| 3236 | TGAAAAATTTCTGGACCTATGTTTGAAGCTCAGCTT.....CCAGTTA | 3193 |
| 663 | erPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMet | 679 |
| | | |
| 3192 | ATTG.....TTTACACTACTTCATGTCTGTATGGTCG | 3158 |
| 680 | AspAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGl | 696 |
| | | |
| 3157 | ACAATAATTTTATTATTTCATTCAAAGCTATGTTTCAAAAAATGAA..... | 3110 |
| 696 | ySerLysIleGluGluAspPhe..... | 703 |
| | | |
| 3109 | AACAAATTAACGTGGATCTGCATCAGCTGTTAAAACCTCTTTACCAG | 3061 |
| 704 |ProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSer | 718 |
| | | |
| 3060 | GGAATCCTGAACCTGTTTCCGCTGCTGCCATTTTAA..... | 3023 |
| 719 | LysLysGlyProProArgGlnAlaLysIle.....AlaIleHisCys | 732 |
| | | |
| 3022 |AGCGTAACAAATTTATGGATATTTTGGAGTACTG | 2988 |
| 732 | sIleHisAla..... | 735 |
| | | |
| 2987 | TATTATTCCAATCCAAATTTTACTAAATTCATCCCTGCATCAATTCAG | 2938 |
| 735 | | 735 |
| 2937 | ATCTCTATTGAAGCTGTTTACCACGCTAGCAGAGCTTCATATTTTGCAGCA | 2888 |


```

MEDLINE 95112788
FEATURES
source
Location/Qualifiers
1. .6449
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="II"
/gene="RIF1"
complement(182..5932)
CDS
complement(182..5932)
/gene="RIF1"
/feature="RIF1"
/feature="ORF YBR275c"
/codon_start=1
/protein_id="CAA85238.1"
/db_xref="GI:536715"
/translation="MSKDFSDKKKHTIDRDOHLRSDHNSNGSSPMWKTNLPP
SPOAHMIQSDLPTRPKRRKLASSDCENKQFDLSAANKNLYPEDTGRUWQPELS
ASNDNSVPYTKSAFSDRIESSPIYRPGSSPKPSPSSKPKGKSIILNRRLSVRTVS
LSYLNQQTQHLKANGNIFTSPYRETRVNRPALEIYWSGEIHGLVDNESVSEKEIEI
GGLGILROESDYVARFEVATFNPIITLTKNNEVDQKFNILVNIESIIEICI
PHIAQDTLLSSEKKNPFVIRLYQVIRFSAIMSNFKIKVLTWRKRPDLNKLKVI
RYWTGALRNENSKIIITAQVSLRDEKFGTFELSNIEIKPIISTETEMEINSHNL
IYKLLIRGLSPVKPLMIETVTSWLPGEVLPRIIIGDEIYSMKILITSVLLELL
KKCLDFVDEHERIYQCIWLSPVCEIPEKFLSKPLASYSANLDKVTIGHLLTQOIK
NYIVKNDKIAMDLWLSGLLYDSGRVYDLSSESNKVFWDNNLCFINHPKTRL
MSIKVRLITICIKISOKNOEKNLSLLRTPFQMTLPYVNDPSAREGIIYHLLG
VYVTAFTSNKLSMDLEFWDHLITPIYEDYVFKYDSIHLQNVLTFLVHLIGKNA
DLALDLILSLKHLPEKDKQTHLESFSSLLIKVTQNNKOTPIFRDFEGAVTSFVVT
FLDLFLRKNDSLVNFNIOISKVIGSQGNWTLDLLKQVIRKARNETSEFLIIEKLEL
DKRTEVTAQNWGSLTFPIQSYSKNENKLNVDLITLTKSLPQNPFLFGLLFPVR
LFTLLHVSMTNPNLNSIPDLNLSLLKLPRLSRASYFAANKIFKCSQQLTVR
NFKMDILLEYCIHSPNLSIPDLNLSLLKLPRLSRASYFAANKIFKCSQQLTVR
WLLGQOQLNQNFSEIENVQNASDSELEKSEIIRELLHMANPIELFSGLLNF
CIKNMADHLDKFCQGNMTSEVLFKISPELLIKLITYKEKPNGLKLAIVIEKIENGDD
YILEELKIIQKIQILEKLEPLVFNPSNOKHKKSTMLRELVLVLTTP
LSRSAAKKFSKSLISLPPNRYQOTIDMNLILDLIKSHNRKFKDKTYNATLTIGK
EPPSGIISQSDSKLEAIPDKSMYIPCEGSENKUSLNQKRVDSQDIQVPTQGMK
WQSSQIISQSDSKLEAIPDKSMYIPCEGSENKUSLNQKRVDSQDIQVPTQGMK
EPPSGIISQSDSKLEAIPDKSMYIPCEGSENKUSLNQKRVDSQDIQVPTQGMK
IDQMKSTQDLNPNVNSDNLKPSKETPTSENKSGKTHETVMDVILVSEQAVSPSS
AVTNRSKTQDLNPNVNSDNLKPSKETPTSENKSGKTHETVMDVILVSEQAVSPSS
DVQENSTKSTANESSILALRNSIKVETNCNENSLNVLDDLOQTITKEDKGVEHVQ
GIQIPSTILLNSKQTEKSKYDLDLSDDEHGTVAQKQVGAIRNKNRNDST
PIQTEESREVMTEGINVRLEDGTCENLNLKGLPKGDKDANINDPVPVEENY
RQEGFLKSMHVAHESKETGLEEPEVADISVLPEIRIPIFNSLRKMQGSKQIEKILKR
LQRNELPPDPPPTMTNTINAQGLDTPVKITGGKKEHIEIQLOQAHTBADGEPLL
GGDGNEDATSEATPSPSLKVHFFSKSRRLVARLGRFTPGDLNGISVEERNLRITELD
FMRLSEYYSNRNDNMN"
misc_feature
3859..3869
/feature="ARS-consensus"
BASE COUNT 1829 a 1213 c 1086 g 2321 t
ORIGIN
alignment_scores:
Quality: 227.00 Length: 1819
Ratio: 0.293 Gaps: 80
Percent Similarity: 42.551 Percent Identity: 17.372
alignment_block:
US-09-512-581-2 x SCYBR275C/rev ..
Align seg 1/1 to reverse of: SCYBR275C from: 1 to: 6449
3 HisSerLysThrArgThrAsnAspGlyLysIleThrTyr..... 15
5386 TATAAGGAACACAGGTAATCTCCGGCTTAGACTATGGTGAGTGG 5337
16 .....ProProGlyValLysGluIleSerAspL 25
5336 TGAGATCCATGGTTTAGTGATAATCAAGAGTGTTCGGAGTTTAAAGAAA 5287
25 ysile.....SerLysGluGluMet 31
5286 TTATTGAGGCGGATGGGCATCTTCGCACAGAATCTGAAGATTACGTC 5237
32 ValArgArgLeuLysMetVal.....Vally 40
5236 GCGAGGAGGTTTGGAGGTATACGCAACCTTTAATAATATTATACCCATCT 5187
40 sThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLysTyrL 57
5186 AACCACTAAATAATGTTAACGAGGTTCATCAGAAATTCATATTCTAATTC 5137
57 euAsnLeuAla.....LeuHisLeuAla 64
5136 TGAATATAGAAAGTATCATTAATAATATGCATACCCGATCTACAATAGCA 5087
65 SerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAl 81
5086 CAGGACACATTTAATAGTAGTAGTGAAAAAATAATCCATTTGTGAATC... 5039
81 acysCysLeuAlaAspPheArgIleTyrAlaProGluAlaProTyrT 98
5038 .....AGACTATATGTTCAATAGTTCGATTCT 5011
98 hrSer..... 99
5010 TCAGTGCTATTATGCTAATTTCAAAATCGTTAAGTGGCTAACGAAAGG 4961
100 ProAsp.....LysLeuLysAspPhe..... 107
4960 CCAGATTGGTTAATAAGTTAAAGTTATTACAGGTGGACCAACCGGTGC 4911
108 .....MetPheIleThrArgGlnLeuLysG 116
4910 GCTAAGAAATGAAATTCACCAAGATAATCATACACGACACAGTCTCTT 4861
116 LysLeuLysAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeu 132
4860 TCTTGAGAGACGAAAAA.....TTTGGTACTTTTCTTGGAGCAAC 4820
133 GluAsnIle...AlaTrpValLysSerTyrAsnIleCysPheGluLeuG 148
4819 GAAGAAATTAACCTATCATAGTACCTTTACCTGAATAATGGAATCAA 4770
148 uAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerVal 165
4769 TAGTCACAACTAATATATGAAAAATTG.....CTAC 4738
165 leAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMet 181
4737 TTAATAAGGGGATTTTGGACCAAGTATCCAAAAATTAATGATTGAGACGGTA 4688
182 SerSer.....IleIleCysGluGlyAs 189
4687 ACTTCTGTGTACCAGGTGAGTACTACCTAGGATTATTATTCGGAGATGA 4638
189 pThrValSerGlnGluLeuAspThrValLeuValAsnLeuValProA 206
4637 AATTTACTCCAGAAATCTCATCACTTCAATAGTTGTTTACTG.... 4592
206 laHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeu 222
4591 .....GAACCTACTATA 4583
223 LysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnG 239
4582 AAAAAATGCTTAGATTTTGTGATGACATGAAGAGATTAC.....CA 4539
239 nValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValP 256
4538 GTGTATCATGCTATCTCCAGTATGCGAAACAATCCCGGAAAAATTTTAT 4489
256 heAspLeuIleLeuGluLeuTyr.....AsnIleAsp..... 266

```


992 lProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValG 1009
 89153
 1009 lnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeu 1025
 89166 AGGATATTGAAATAATAAAAAAT
 1026 GluIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIleArgLy 1042
 89189
 1042 smetValGluAsnIleLys...GlnThrLysAspAlaGlnGlyPro... 1056
 89227 ATTAAATAGCATATAAATAACATATGAAATAATACAGCAAAATAT 89276
 1057
 89277 TAATAATTGATGATGCCAAAAAGAAAGTAAAG
 1071 ValAlaMetAsnIleIleMetSer...LysSerThrThrTyrSerLeuG 1086
 89312 ATAACAGATAATATTAACAAGCGCTTTTAATGAAATACAGAAAAATTATA 89361
 1086 userProLysAspProValLeuProAlaAArgPheThrGlnProAspL 1103
 89362 TAATGAAATAATGCGGTAAATAAATCTGCA
 1103 ysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPhe 1119
 89394 AAAATATTGCGATAAAGCTACTTATTAAATAATGAATTAGATAAATTT 89443
 1120 PheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLy 1136
 89444 TTATTG.....AAATGAATCAATTTATTAAATGCTATAATAATA 89481
 1136 sProLeuSerSerAlaGly.....LysGlnSerGlnT 1147
 89482 TGATATAAAGATCTGCTGATGAAATAATTAATTAAGAAGAAGAAG 89531
 1147 hrLysSerSerArgMetGluThrValSerAsnAlaSerSerSerAsn 1163
 89532 AAAGAAAGAAAGA...GAAAGATTGGAAGAAAGCGAAACAAAGAAGAA 89578
 1164 ProSerSerProGlyArgGlyIleLysGlyArgLeuAspSerSerGluMetAs 1180
 89579 AGAAAGAGAGAGAAAGAATAAGAAAAAGAAAAACAAAGAAAGAGAGACT 89628
 1180 pHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysL 1197
 89629 GGAAGAGAGAGAAACAGAACCACTAAAGAAAGAGCAATTAAGAAAGAAAG 89678
 1197 ysSerAspLysArgAsp.....AspSerAspLeuValArgSer 1209
 89679 AGCAAGAAAGACAGAACACAAACAAAGAAAGAGCAATTAAGAAAGAA 89728
 1210 GluLeuGluLysProArgGlyArgGlyLysThrProValThrGluGlnG 1226
 89729 GAACAAGAA.....CGACTACAAAGAAAGAAAGAAATTAAGAAAGACAA 89772
 1226 u...GluLysLeuGlyMetAspLeuThrLysLeuValGlnGlnL 1242
 89773 CCAAGAAAGCTGGAAGAGAGAGAAACAGAACCACTACAAAGAAAGAAAG 89822
 1242 ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258
 89823 AATTAAGAAAAAGAGCAGCAAGAAACAAACAAAGAAAGAAATCCAGAA 89872
 1259 SerAspGluGlnGlnTrpProGlu.....GluLysArgLeuLysG 1272
 89873 TTGAAGAGCAAAAAAGCGCTGAAATAATAAATGAAGCAATTTGGTAAAGGG 89922

1272 uASP...IleLeuGluAsnGluAspGluGlnAsnSerProProLysLysG 1288
 89923 GGATAAAATACTAGAGGAAGTATGATCAGAGAAATATGCAATTAAGCAA. 89971
 1288 lyLysArgGlyArgProProLysProLeuGlyGlyThrProLysGlu 1304
 89971
 1305 GluProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyPr 1321
 89972 ...CCTACGTTAGTATGGATACTAATAATAATAGTCCAATTAAGTAACAG 90018
 1321 oProAlaProGluGluGluGluGluGluGluGluGluGluGluGluGlu 1338
 90019 TGAATTAACAGAAACGCGATGATATTGATAACAGTCAAAATATACATACTA 90068
 1338 luGln.....LysSerLysSerLysGlnHisArgValSerArg 1350
 90069 GTCATATGATGATGACATCGAAAGTACACAAACTAGTCTAGAGAAGTAACACC 90118
 1351 ArgAlaGlnGlnArgAlaGluSerProGlu 1360
 90119 CATGGCAACAAATCAGTGATATTGTTCAA 90148
 seq_name: gb_pl4:SCYBR275C
 seq_documentation_block: 6449 bp DNA PLN 11-AUG-1997
 LOCUS SCYBR275C
 DEFINITION S.cerevisiae chromosome II reading frame ORF YBR275C.
 ACCESSION Z36144 Y13134
 VERSION Z36144.1 GI:536714
 KEYWORDS
 SOURCE Saccharomyces cerevisiae.
 ORGANISM
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE
 1 (bases 1 to 2920)
 AUTHORS Aigle,M., Baclet,M.C., Barthe,C., Biteau,N., Crouzet,M. and Doignon,F.
 JOURNAL Unpublished
 REFERENCE 2 (bases 2644 to 6449)
 AUTHORS Brandt,T., Christiansen,C., Holmstroem,K. and Kallese,T.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 6449)
 AUTHORS MIPS.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-1994) Data collected by MIPS on behalf of the European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
 REFERENCE 4 (bases 1 to 6449)
 AUTHORS Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C., Barthe,C., Baur,A., Becam,A.M., Biteau,N., Boles,E., Brandt,T., Brendel,M., Brueckner,M., Bussefau,F., Christiansen,C., Contreras,R., Crouzet,M., Czepluch,C., Demolis,N., Delaveau,T., Doignon,F., Domdey,H., Duesterhus,S., Dujoon,B., El Bakoury,M., Entian,K.D., Feuermann,M., Fiers,W., Fobo,G.M., Fritz,C., Gassenhuber,H., Glansdorff,N., Goffeau,A., Grivell,L.A., de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K., Jacq,C., Jacquet,M., Jauniaux,J.C., Jonniaux,J.L., Kallese,T., Kiesau,P., Kirchraht,L., Koetter,P., Korol,S., Liebl,S., Logghe,M., Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallet,L., Mannhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S., Nasr,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E., Pohl,F.M., Pohl,T.M., Potier,S., Profit,M., Purnelle,B., Ramezani Rad,M., Rieger,M., Rose,M., Schaeff-Gerstenschlaeger,I., Scherens,B., Schwarzlose,C., Skala,J., Slonimski,P., Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R., Urrestarazu,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A., Vetter,I., Vierdeels,F., Vissers,S., Wagner,G., de Wergifosse,P., Wolke,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Klein,K. Complete DNA sequence of yeast chromosome II
 EMBO J. 13 (24), 5795-5809 (1994)

302 AlAlYAspSerGluLeuAlaSerGlnAsnLysProLeuTrp..... 315
||| : : : ||| : : : ||| : : : ||| : : : |||
86702 AAAAAAAAAAATTTCGAATCTAAATAACAACATCATTAACACTACTTAAAT 86751

316GlnCysTyrrLeuGlyArgPheAsnAspIleHisV 327
: : : ||| : : : ||| : : : ||| : : : ||| : : : |||
86752 GGATTCATTCAGACTGTATGTTTATTAAGTAATATAAATGATTATATA 86801

327 alProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsn 343
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
86802 TAAATGAATAATTTGAA.....AAA 86821

344 HisProAspLeuAlaLysAspLeuThrGluTyrrLeuLysValArgSerH1 360
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
86822 CATCAAAATATATTGAATGAAATATATAATGGATTATTAATGAA..... 86863

360 sAspproGluGluAlaIleArgHisAspValIleValSerIleValThra 377
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
86864TCATATAATATTATTAATACAAAAATGACTG 86894

377 laAlaLysLysAspIleLeuLeuValAsnAspHisLeu..LeuAsnPhe 392
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
86895 AA.....ATTATAATGATAATTTAGATTATAATGAA 86926

393 ValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMe 409
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
86927 ATAAAAGAA.....ATTAAAGAAGTAGCAC 86952

409 tMetGlyLeuAlaGlnIleTyrrLysLysTyr.....AlaL 421
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
86953 AACAGAATATGATAACCTTAATAAAAAAGTTGATCAATPAAAAAATTATT 87002

421 euGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLys 437
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
87003 TGAATAATATTAAAGAACAGAGGACATCGATTAAITGATTATATAAA 87052

438 AspLysLeuLeuHisIleTyrr.....TyrrGlnAsnSerIl 449
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
87053 GAAAAATATTAACTTATATAAAATGTTCCAGAACACAAATATAAT 87102

449 cAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrrMetValProH 466
||||| : : : ||| : : : ||| : : : ||| : : : |||
87103 AGATGATTCTTATATTATATACAGTTAAAAAACAGTAT..... 87142

466 iAsnLeuGluThrThrGluArgMetLysCysLeuTyrrTyrrLeuTyrrAla 482
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
87143ATTAAACATTTCAGAGTGGAATTTTTA..... 87172

483 ThrLeuAsp...LeuAsnAlaVal.....LysAlaLeuAsnGl 494
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
87173 ...TTAGATTCTTGAACACAAATAGACAAAAAAATAAATCAGTAGCAA 87219

494 uMetTrpLysCys.....GlnAsnLeuLeuArgHisG 505
: : : ||| : : : ||| : : : ||| : : : ||| : : : |||
87220 TCTAGAAATTTGCTACTAATAAGAAGATATAAAAAATTTACTTAAACATG 87269

505 lnValLys..... 507
: : : ||| : : : ||| : : : ||| : : : ||| : : : |||
87270 TTATAAAGTTGGCAATTTTTCAGGTATTATTGTAATGCTGTGATACAAAT 87319

508AspLeuLeuAspLeu 512
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
87320 ACGGNAATAACTCCAGAAAAATCCITTAGAAGTAATGATTATTATAATTT 87369

512 uIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysV 529
| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
87370 A.....CAATTATATTTCGAAGAA 87389

529 alMetValIleThrArgAsnLeuProAspProGlyLysAlaGln..... 543
||||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
87390 AACATGAAATACATCAACATTTGGAATATGATCTCGATTTAGAGTTAGAT 87439

544AspPheMetLysLysPheTh 550


```

723 .....ProArgGlnAlaLysTyrAlaIleHisCysIle 733
      ||||| ||| :|||
8537 TGAGATCCAAGCGTTTCACAGACACGGA...GTTCTCCAC..... 8497
734 HisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPr 750
      ||||| :||| :|||
8496 .....AAGGAGCCAAAGACGGTCGAGAGTTCGAACAC 8465
750 oLeuHisLys.....SerLeuAspProSerAsn.....LeuG 761
      :|||:||||| :|||
8464 TATTCGGAAGAATGATTCATTAAGGCCAAATAATCACAGGATCAATPAC 8415
761 luHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuAla 777
      ||||| ||| :|||
8414 AACAGCCAATAGATGCCAAATTTCT..... 8389
778 ProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThr.....Ph 792
      |||:||||| :|||
8388 ...GACATGTTCTTGAAGCCTTGGAGCCCTTGGAGATCGAAGAAGGG 8342
792 eIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysThrT 809
      ||| :||| :|||
8341 AATTCGTTACGAAAGAAAGATATAGAAAATGTGAAGCAAAAATACTA 8292
809 hrLysLeu.....TrpVal..... 813
      ||||| |||
8291 CCAACTTCAGCTATGCCATGGAACAAATAATGCTTCGAAAACCTTGTCAC 8242
813 ..... 813
8241 TATAATGTTAAACATCAAGCGGTACGATTGAGCGATTTTAAGAAGATTAG 8192
814 ProAspGluValSerProGluThrMetValLysIleGlnAlaIleL 830
      ||| :||| :|||
8191 ACCAAGACGACGAAGATATGGAACCC.....AAGAGAAATCTATTG 8148
830 yMetMetValArgTrpLeuLeuGlyMetLys..... 840
      :|||:||||| :|||
8147 AGATTGGTGAACACCATCCAAACCTCAAGTCGTGTTGAAACGTGGG 8098
841 .....AsnAsnHisSerLysSerGly..... 847
      |||:||||| :|||
8097 CAAGGACATCGTTCTATGAATTTAACTCTCAATCTCCTGATGCAAAAC 8048
848 .....ThrSerThrLeuArgLeuLeuThrThrIleLeuH 859
      ||| ||||| :|||
8047 TTTAGTGGTGGATCCAAACAAACACTGAAGCTTGCAAGACAAAGTGCAC 7998
859 isSer..... 860
      |||||
7997 ATTCGTGGCTGGAACCTCACTCCAAAAAACTCTTTTACTGTTCCGTCG 7948
861 .....AspGlyAsp.....LeuThrG1 866
      |||:||||| :|||
7947 CTCTAGAATTGGACGATTGTCGCCCTCTCAAGAATAACATGGCTGACGAA 7898
866 uGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaG 883
      ||| ||||| :|||
7897 CACAAGAAAAAATTCGAAGGGTGAAGAA.....AAGTTAAGGAAA 7857
883 lySerAlaIleValLysLeuAlaGlnGluProCystYrHis...GluI1 898
      :|||:||||| :|||
7856 ACCTGGACCTGTTCAACAAGCTCCCGTCCCGTCTCCACCAAGAGAG 7807
898 eIleThrLeuGluGlnTrpGlnLeuCysAlaLeuAlaIleAsnAspGluC 915
      |||:||||| :|||
7806 GAAACTCCTGAACCAACACCT..... 7785
915 yTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSer 931
      |||:||||| :|||
7784 .....CGCCGCTGTTCTCTTCCTTCGACCATTCCTCGGCTGAAT 7746
932 Arg.....Le 933

```

```

7745 CGAAGAAAAAGTACCGTCGAGAAAGTTGACACTCTCTTCACTCTGACT 7696
      |||
933 uArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 950
      |||:||||| :|||
7695 CCACCTCGCGCTTCGGTTCCTG..... 7674
950 roValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
      ||| :|||:||||| :|||
7673 .....CTCCACCACCAAAATGCGCAGCGATCAACAGT 7641
967 AsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLe 983
      :|||:||||| :|||
7640 GGAGGCGAGCGATGATGGCTCGAA..... 7617
983 uLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuA 1000
      :|||:||||| :|||
7616 .....GATATATGGGTGCCAGCAAGCGCTGTTTGTGTA 7583
1000 laHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
      :|||:||||| :|||
7582 GCCACCATCGGCTCTTTCGAAAACCTGAATCATTTGAAGGAAGCTTCAAACA 7533
1017 ValLysGluCysLeu.....TrpPheVa 1024
      |||:||||| :|||
7532 TTCAAGAGTCTGCTTCGACGACTAGGTACACCAGCGTTGGTGGTTTCA 7483
1024 lLeu.....GluIleLeuMetAlaLysA 1032
      :||| :|||
7482 CATTCCTACTCCAGAACCTGTACAACTAACAGAATGGCGCAGCTAAAAA 7433
1032 snGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLys 1048
      |||:||||| :|||
7432 ATCCGAACACAGAAGAAAT.....ATAAACACCAAGTGGCAAAACCAAGG 7389
1049 Gln.....ThrLysAspAlaGlnGly..ProAsp 1057
      ||| :|||
7388 CAACAGCACCCCTCCGATCACTCCAAAAGCTGCATCAATCAGCGCTGCT 7339
1058 AspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAs 1074
      |||:||||| :|||
7338 ACTCAAAAGTGCAGCAATCCGAATTCGACAAAAGATGAAGCTTCCAGAAA 7289
1074 nIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAsp 1091
      ||| :|||
7288 TGTTCATCTCT.....TCTGGAAAGTCTCAGAAACAGG 7254
1091 roValLeuProAlaArgPheThrGlnProAspLysAsnPheSerAsn 1107
      :||| :|||
7253 CAGACGTCCAATCTCGG..... 7237
1108 ThrLysAsnTyrLeuProGluMetLysSerPhePheThrProGlyLy 1124
      |||:||||| :|||
7236 .....CCTAAATACAAAGTGTA..... 7219
1124 sProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerA 1141
      :||| :|||
7218 ....CAGGAAAAACGCTCTGCGCAAAAGTGTCAAACCGCTTAGATCCAA 7173
1141 laGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal..... 1155
      ||| :|||:||||| :|||
7172 AAGTCTCTAAAGGACCAAGAACCAAAAGATGAGAAAGTGGTGCCT 7123
1156 .....SerAsnAlaSerSerSerSerAsnProSe 1165
      :|||:||||| :|||
7122 AGTACCAACGCTGACTTTGGATGATCAAGCTTCATCTTCTTCGAATCAGCA 7073
1165 rSerProGlyArgIleLys..... 1171
      :|||:||||| :|||
7072 AGGACCAAGGATCCGCGAAACCAACATCTGCATCAATTTTCACTGCATTAT 7033
1172 ..GlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyr 1187
      ||||| :|||

```


gene
CDS
complement(join(19458..19504,20357..20416,20525..20626,20921..21179,21299..21441,21986..22258,22320..22405,22461..22545,22638..22736,22850..22948,23006..23065,23373..23579,23628..23771,24167..24309,24353..24401,24457..24506,24632..24740,24814..24906,25021..25079,25127..25241,25311..25490))
/gene="F39C12.2"
/name="F39C12.2"
complement(join(19458..19504,20357..20416,20525..20626,20921..21179,21299..21441,21986..22258,22320..22405,22461..22545,22638..22736,22850..22948,23006..23065,23373..23579,23628..23771,24167..24309,24353..24401,24457..24506,24632..24740,24814..24906,25021..25079,25127..25241,25311..25490))
/gene="F39C12.2"
/note="similar to alpha-adducins; coded for by C. elegans cDNA yk26f2.3; coded for by C. elegans cDNA yk26f2.5"
/codon_start=1
/db_xref="GI:2736400"
/translation="MKQENKTSKGVHPFAPFEKHNSENHRTFRILLFLFNSLYLT
LFHFHCKVMFTFLNFMKGRKEKPERYPDPPEYIKDLQRPVAVIKVSAAPTLFF
LFFQLOLMKPPIIAEDLSEMERKRVQOILSKSFCEHELEVRQECDTARTDPDH
LOVLQKLSDLTVQGNMSEFNLTENFPIRLAMLPQDTRIPMHKQNCNTFLNLF
CGNTAIDLKNEKYSKRAERIOENKACFLRLADLFOWSQGIHNEISVYRTNDENTF
LNNPCLLYHEITAATIVKIDENKILDCGLTLKAGVNQPAFLHLSAIIYKAHPWRCIL

alignment_scores:
Quality: 227.50 Length: 1923
Ratio: 0.319 Gaps: 82
Percent Similarity: 37.077 Percent Identity: 17.733
alignment_block:
US-09-512-581-2 x CELF39C12/rev ..
Align seg 1/1 to reverse of: CELF39C12 from: 1 to: 40233
105 AspilePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLy 121
11283 GATGCTTTATTTTCAGGAACATGATGAGGCTTCAAAACATTACA 11234
121 sserProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpV 138
11233 AGCGCATCAACTGCTGGATTTTCCAT.....CGTCAATTTAACTGGG 11190
138 al.....Lys 139
11189 TGACTTGGAGAATGATGAAGATAAATCTGAACGAGTTTTTGTCTCTGAA 11140
140 SerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrG1 156
11139 GAATTTGATGGCATTTCACCGTGACGCTGATGATAAGTTTGTCTCGGA 11090
156 nLeu.....TyrArgT 160

11089 AATACTTTGTATGAAGACAGGTCGCCAATCGTCTCGCTTCGGCCTTCGGA 11040
160 hrLeuPheSerValIleAsnAsn..... 167
11039 GATTTTTCGCTTCCTTAAACCGCGAGCTGAGCTTCCTGTAGTGAGAG 10990
168 GlyHisAsnGlnLysValHisMetHisMetValAspLeu..... 180
10899 GCAAGAATCAAAAAGTCTCGGCGCTGCAAAAGATTTCACAGTCGCTCTA 10940
181MetSerSerIleI 185
10939 ACTGCAATACCTCACACCAITTTGAGCGCAGCTGCACCTATCGCACTATTTA 10890
185 leCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeuVal 201
10889 TATGCCGCAACACCAACCACTTTGATTTTAAAGTTAGACTGCAGAAATTGAG 10840
202 AsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAl 218
10839 GCAATGACA.....AAATCAG 10823
218 aLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrT 235
10822 GCGGCGCTTGTCTGGAACCTCGAACCTCCCATCGAAGCTCCTATAAAC 10773
235 hrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeu 251
10772 CTTTATTTTTCAGATGATGACATGATGCGAGACACCATGAATACCATCAGC 10723
252 SerGluHisValPheAspLeuIle..... 259
10722 CAGGCGCTGAAATGGACATGATGACCATGATCAGCATCAGCCCGCGGAAT 10673
260LeuGluL 262
10672 GAGAATTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10623
262 eutyAsnIleAspSerHisLeuLeu..... 270
10622 AGCAGAATGAAAGGTTTCATCTATCTATCGCAAGAAAGTGGCGGAGAGAGCA 10573
271LeuSerValLeuProGlnLeu.....GluPh 279
10572 CGTATTCGTCTGAGAGATTTTCGTGAAGCCGCTTTTGTGTAATGAGTT 10523
279 eLysLeuLysSerAsnAsp..... 285
10522 GAAGAAAAAGGCAAGGATAAGGAAGTGGCTGCAACAATTCCTACTGAAA 10473
286AsnGluGluArgLeuGlnValValLysLeuLeu 296
10472 ATGATGATGCCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10423
297 AlaLysMet..... 299
10422 AAAAGTTGCCAATTCCTGAGTTTATGAGATCCGAAAAAGAGAGATGCC 10373
300PheGlyAlaLysAspSerGlu.....LeuAlaSerGlnAsnLysP 313
10372 AAACCTTGCACCAACCAAAATCAGAGGATTGGCTTGAATCATCTGATCT. 10324
313 roLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIle 329
10323GAAGACGAACTTCCGCCGATC 10303
330 ArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAs 346
10302 AAATCTCAATT..... 10291
346 pLeuAlaLysAspLeuThrGluTyrLeu...LysValArgSerHisAspP 362
.....


```

REFERENCE
AUTHORS
1 (bases 1 to 40233)
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
Saunders,D., Shownkeen,R., Smalton,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
MEDLINE
REFERENCE
1 Nature 368 (6466), 32-38 (1994)
94150718
2 (bases 1 to 40233)
Chissoe,S. and Sansone,J.
TITLE
The sequence of C. elegans cosmid F39C12
JOURNAL
REFERENCE
1 Unpublished (1997)
3 (bases 1 to 40233)
Waterston,R.
TITLE
Direct Submission
JOURNAL
REFERENCE
1 Submitted (18-DEC-1997) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:
1. 40233
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/complement(212).18270)
/gene="F39C12.1"
complement(join(212).469,519..665,889..1137,1192..1327,
1377..1546,1675..1923,1983..2286,2331..2431,2478..2921,
2964..3128,3178..40758,11107..11265,11315..11701,
11775..12658,12722..13094,13141..13596,13653..13961,
14015..14191,14239..15440,15491..15577,15622..15793,
15845..16066,16113..16227,16276..16725,16780..16875,
17643..17759,17836..18002,18059..18143,18215..18270))

The 5' cosmid is R08E3, 2000 bp overlap:3' cosmid is T07F12, 200 bp
overlap. Actual start of this cosmid is at base position 1997 of
CELF39C12; actual end is at 5117 of CELT07F12

NOTES:
Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1. 40233
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/complement(212).18270)
/gene="F39C12.1"
complement(join(212).469,519..665,889..1137,1192..1327,
1377..1546,1675..1923,1983..2286,2331..2431,2478..2921,
2964..3128,3178..40758,11107..11265,11315..11701,
11775..12658,12722..13094,13141..13596,13653..13961,
14015..14191,14239..15440,15491..15577,15622..15793,
15845..16066,16113..16227,16276..16725,16780..16875,
17643..17759,17836..18002,18059..18143,18215..18270))

/gene="F39C12.1"
/note="contains a short region of similarity to troponin T
cardiac isoform; coded for by C. elegans cDNA yk38l1.5;
coded for by C. elegans cDNA yk31c2.5; coded for by C.
elegans cDNA yk44a1.3; coded for by C. elegans cDNA
yk44a1.5"
/codon_start=1
/protein_id="AAB94194.1"
/db_xref="GI:2736399"
/translacion="MTDPDGIQDIFIKFICPNEEFGLKREMFISERIIVEYDPTD
RLTSKGNLCSLGEITVIGRLDGLKMLAVNHTDSSVNTVTVDLWPKPAENIAH
YDSVSPNTDNIWHGFFLLAVGDESLRVMDRCRPAFNAVFDGGVQKALCSTRP
LEIACWNSNVLVDVVRPEVPCAIYDGLKESNIDLMKFNPIINELFVHTNSHRV
FAMNMPHTLIGRPISSEKPYKQVITFKHIDRGCTATHTMLRREKMPDEYVPL
LGYYGYEAAKNQDQVDFPTLWTVPAEVDMKFLAKIQDVIDEYSLYKAEVRSAE
RDKHKPKILYQAKGVESFDLPFGVNGIMCFORTFTYKIVRSPKHGVSVPYAG
SDSVFVYRGLQIOEMIPAQTLTMCMNRLAGWNAVQVSESLKELAKITESSVNA
KWLHNSVIOYGGKEFGNIGVNDLEDVLPDGRPLPEVEIPPEYKMDNVAITEET
OMVAFPLFLPDPRDLSDSESETEDHNAPEGNCEDSDPEGLSPSELSNMRT
LCCHAKIRYNNKQSDRYKMYASMGNSFAIGSKRIKDKFFKLGWSITKPKQILGHR
LKIDTAFKKSILDDLKPKFORFKRFPKLVLPKLTJTDSTSVFMKPEYIKREBAKER
EMREKAAPIQATIGLSMPKSRKEPKQTPVSDNGEEDPEHAFWIRAAIOADNN
SSDDEIVSPQATPTPPAPTPTPHVQALSVRNEKDDFEVETEDAVEGTKKRMVY
YQPATOPGLGRYKPDRLRESTFIRLEHLHMYGRRRHERDILYGFEEENDKQSSV
KERTDLKFSLNHHMKVGMOTFYDKNFFFSNCKSYNOILKKAIDOMNRCGLGFSHET
YFKSVRHRMLCGVNRNVMDKEAIEEARYLRKDISIENSTIKLYNHFMTFYQAQLP
YKLDHTIFMENFEGDPTFSQAECPIEKKQTAKYDENDKFMFLGLRLCFCFETSPN
FMHEINRVPKPKPNGLYLEDPEHLSAQSPKPLVRRNSCPLGLAISTENSRV
CNEFFHEIYRIIAYRDMCHOREHGVNHYGVSEYLPBLTAGVLOQFNTHSVNGDYK
VDDLNTNFKLEWDSMISDIRETRGCRFMRRVVPKPEFQPLQEGIDFLDYIIL
EVMRAAKNLHYSIIFNDILRTWPGIQLDEGRDMFTLCFPFRVPAEQLSEIDIF
SKPLCPTKAVEFYSDFFTEVLLIGRDSKKNRIPRPPVYNEHNIINPEGLILKPAFY
GFFKYDRESYNEEQSLEMKNRRELRAIKRIITPRQPIVEETEOLATEENLWE
EDHLLTDSYNCENPHTREARKISVEEEVEDEPKLGSVLTNLFEQDLVKORLGD
IDFADTDESEGEDEVKNDSDNDISDFWKIFKVTDEAGVNDGSDSDVAVNTKIG
IETDPRGFKTLOGDLQAEFHEPNTWRMDEKSERVSEEDVAFADDAIDMM
RDYMTISQDGMMDMTYDQRTGMRIRRWISYAAATRVQOKNEHYKAAKAREEA
RILSGELSVPSLLNELKKAKDEVAATATENDDAKKEERKAYLKKKLAIPPEE
IEDPKAMPNAPPKSDWLESSEDELPPIKSKILSKEMKEFIDIRVHMEQPKS
FEGIATVYKGLFEKRPILKNDIVLVTTRKNRRKPPFPRIQRIPLDESMLNTATVSF
LPVQVNRPDMPKMTVPYNSORLALTAQETKERALPPPTPTTKLKYTLTKWDVE
KAFYHRRKVSQITLKAAROLRHLFKLYHSKPSHPTKAWTHVRQIGLKKSVY
PKYOKRVYELKEDHPTNSELGIIERKKELKREDPAVLMSSTRSTLKEPETSFKV
REISVGPAIPKADVQKLTSLKIDSQTEPKTESIKQRHPIFHELEDKSIINFS
QGIPIASVPEPSCQPKRQKQKHEKITNSNSNSIDSGMKMRSSSKQPSVAVRS
KSTMIPHSRLRKQMPACRSFVDSKSSYKALKIARKSIOKRSKVKYTPDDSK
PEVPSDQPKPSQLETSVALPSKAKVSTKSEKDLPLKSTELQELERKFLKNVHNR
RSSRPGIHHORPLHPLDTHETDRKSRAASQFRFOHMRKRETRORVLLHKEPR
RSSRLNTIRKNDSLGKNHMKYEQPIDAKISDMFLKPMKLRBSKKGSRKYYKN
VKANTKLOPMKQMLKADYNVIRKRLSDFKIRPKQEDMETREKSEIIVY
KPSNLSKLRGGRHSMKPNPSQDKTSAGSNKTLKLVTSLSHLSLQTSPPKTS
FTVPSLELDDSDSLKTLWTNTNKSQKKEKKVKNTPAQVAPVPSPPREPTPEPT
PSPVRLRLRLSAESEEBKPVKEVDTPSLTPPAASVAPPKPPKPAQMQPMALRMAA
YMGASKPVFKVPSALSSTESKEASNIQESATRVQPALVSHSTPEVPLTRMAA
AKKSEQEKYTPVAKPKATAPPITPKAASIRPATOKLHESELTKSEASRVSSSGK
SQKQADVQSRPKLQSVQEKRSVPKVPPLDKPGQKEGSKKNEKVPSTNVLDDQA
SSNSQPGGSAKPSASITAFGRKKRKMSTSAESSIPIPSPENQATSETNR
SSPLPRIKKPTQLIKRSTSDVPKPVVSGAIVKFDGKPIISNPSYEGPKMARR
PKKADSAADSGDLSDEGSHGELKDEENSKQKAGTKVRFQIYEPPEPAPARR
VEIPKVPETPEISFESLAQOAKOTLRKEASKILEHPVTEQTSPTABVSPTKSG
SKLSFGLSGEGVTKPKNTAQTSLVLESTSSPELLTPSNPSTSTASSSKQPK
SLSEKITAQMTQTRGRINKLSHRRQLKIGDPSLQSGDSSLRPAQSPKPKPO
LEVPGVQVTNPPSQRYITRLGHEWAPPKPAKTPLSIRVPSQSGVSTSTPAIDLSAKE
VLAPIETKTKMRKRPKPMETTLVLDGTBEHFKLTAETLQSLRQKSKRNNKKRT
VLAISVIGDENDSDTKCMQHNKLLLEERDLKLSLTQLQDEKMITASAVNFPRTDSS
STKHELLRNPKFETTKLKLRLANNRQKTRMSGIHSQPKRDVAISAIROVKVRF
QSTGCKMMNPISGFKPSILKPKPATWLFRLDSEGNKLYMKLGLPSDQKNDKEK

```


* 24376 27231: contig of 2856 bp in length
* 27232 27331: gap of 100 bp
* 27332 31148: contig of 3817 bp in length
* 31149 31248: gap of 100 bp
* 31249 33450: contig of 2202 bp in length
* 33451 33550: gap of 100 bp
* 33551 36754: contig of 3204 bp in length
* 36755 36854: gap of 100 bp
* 36855 39988: contig of 3134 bp in length
* 39989 40088: gap of 100 bp
* 40089 42367: contig of 2279 bp in length
* 42368 42467: gap of 100 bp
* 42468 45775: contig of 3308 bp in length
* 45776 45875: gap of 100 bp
* 45876 49987: contig of 4112 bp in length
* 49988 50087: gap of 100 bp
* 50088 53225: contig of 3138 bp in length
* 53226 53325: gap of 100 bp
* 53326 56622: contig of 3297 bp in length
* 56623 56722: gap of 100 bp
* 56723 59393: contig of 2671 bp in length
* 59394 59493: gap of 100 bp
* 59494 63369: contig of 3876 bp in length
* 63370 63469: gap of 100 bp
* 63470 68148: contig of 4679 bp in length
* 68149 68248: gap of 100 bp
* 68249 72051: contig of 3803 bp in length
* 72052 72151: gap of 100 bp
* 72152 78079: contig of 5928 bp in length
* 78080 78179: gap of 100 bp
* 78180 82389: contig of 4210 bp in length
* 82390 82489: gap of 100 bp
* 82490 87187: contig of 4698 bp in length
* 87188 87287: gap of 100 bp
* 87288 93791: contig of 6504 bp in length
* 93792 93891: gap of 100 bp
* 93892 99177: contig of 5286 bp in length
* 99178 99277: gap of 100 bp
* 99278 104756: contig of 5479 bp in length
* 104757 104856: gap of 100 bp
* 104857 112844: contig of 7788 bp in length
* 112645 112744: gap of 100 bp
* 112745 119589: contig of 6845 bp in length
* 119590 119689: gap of 100 bp
* 119690 130279: contig of 10590 bp in length
* 130280 130379: gap of 100 bp
* 130380 144451: contig of 14072 bp in length
* 144452 144551: gap of 100 bp
* 144552 160962: contig of 16411 bp in length.

FEATURES

Source

1. .160962
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone_lib="RP11-343C9"
1. .1187
/note="assembly_fragment"
1288. .2962
/note="assembly_fragment"
3063. .4458
/note="assembly_fragment"
4559. .5583
/note="assembly_fragment"
clone_end:T7
vector_side:right
5684. .7420
/note="assembly_fragment"
7521. .9277
/note="assembly_fragment"
9378. .11138
/note="assembly_fragment"

misc_feature 11239. .13265
/note="assembly_fragment"
misc_feature 13366. .15440
/note="assembly_fragment"
misc_feature 15541. .17308
/note="assembly_fragment"
misc_feature 17409. .19894
/note="assembly_fragment"
misc_feature 19995. .21852
/note="assembly_fragment"
misc_feature 21953. .24275
/note="assembly_fragment"
misc_feature 24376. .27231
/note="assembly_fragment"
misc_feature 27332. .31148
/note="assembly_fragment"
misc_feature 31249. .33450
/note="assembly_fragment"
misc_feature 33551. .36754
/note="assembly_fragment"
misc_feature 36855. .39988
/note="assembly_fragment"
misc_feature 40089. .42367
/note="assembly_fragment"
misc_feature 42468. .45775
/note="assembly_fragment"
misc_feature 45876. .49987
/note="assembly_fragment"
misc_feature 50088. .53225
/note="assembly_fragment"
misc_feature 53326. .56622
/note="assembly_fragment"
misc_feature 56723. .59393
/note="assembly_fragment"

alignment_scores:

Quality: 228.00 Length: 58
Ratio: 4.145 Gaps: 1
Percent Similarity: 94.828 Percent Identity: 79.310

alignment_block:

US-09-512-581-2 x AC022862/rev ..

Align seg 1/1 to reverse of: AC022862 from: 1 to: 160962

351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
:::|||||
123366 ATTACAGATATTTAAAGGTTAGATCATGATCCAGAGAAGCTATTTCG 123317

367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
|||||
123316 TCATGATGTCTATTGTTACTATAATACAGCTGCCAAGAGGACCTGCCT 123267

384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
|||||
123266 TAGTAATGATCAGCTGCTTGGCTTTGTAAGGGAAGAACTGGATAAA 123217

401 ArgTrpArgValArgLysGluAla 408
|||||

123216CGGGTAAATCTGAGGCA 123199

seq_name: gb_in3:CELF39C12

seq_documentation_block:

LOCUS CELF39C12 40233 bp DNA INV 01-JAN-1998
DEFINITION Caenorhabditis elegans cosmid F39C12.
ACCESSION AF039043
VERSION AF039043.1 GI:2736398
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans strain=Bristol N2.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.


```

644 nAlaIleArg.....AlaGlyLeuGluLeuLeuLysValLeuS 657
   |||  ::  ::::: |||||
12271 AGCGAAAACTATTAGAGAAATCAAGCGCTAGATTGTTGAAA..... 12229
657 erPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeu 673
   |||  ::  ::::: |||||
12228 .....AACGCTAAACCGATGAAGACGG 12205
674 LeuAlaCysLeuLys.....MetAspAspGluLysValAl 685
   |||||  ::  ::::: |||||
12204 AAGAGTGCTTGAAGAACTCCCTAAAGACTTACAGAAAAAGGTTTAGC 12155
685 aGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluA 702
   |||  ::  ::::: |||||
12154 CAAGAGAGTGTTAGGTTTATTGGATTGCGGTATCAAAAGCCAAAACG 12105
702 sp.....PheProHisIleArgSer 708
   ::  ::  ::::: |||||
12104 AAGCTGAAAGAAAGAAATGCGAGAAATGCTCACCCCTGAAGCGAGAAAG 12055
709 AlaLeuLeuProValLeuHisLysSerLysLysGlyProProArgG1 725
   |||  ::  ::::: |||||
12054 CTATTA.....GAAGAGCTAAGAAG.....AG 12032
725 nAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluT 742
   ::::: |||||
12031 CGTTAAGCTTACAAAGACTGCGTTTAAAGAGCTAGGAATCAAAAGAGA 11982
742 hrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSer 758
   |||
11981 AACAA..... 11977
759 AsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
   ::  |||  ::::: |||||
11976 GAATGCGAGAAATTAACACGCT..... 11953
775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrP 792
   ::::: |||||
11952 .....GAAGCGAGAACTATTAGAGAAATCTAAAGAAAGGTTAAGGCTT 11907
792 helLeValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThr 808
   ::::: |||||
11906 ATTTGGATTGCGTATCAAAAGCCAAAACGAAGCTGAAGAAAGAAATGC 11857
809 ThrLysLeuTrpValProAsp.....GluGluValSerProGluThrMetVa 824
   |||||  ::  ::::: |||||
11856 GAGAAATGTCTACCCCTGAAGCGAGAAAGCTATTAGAAGAGCTAAAGA 11807
824 lLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysA 841
   ::::: |||||
11806 GAGTGTTAAGCTTACAAAGACTCGCTA.....TCAAGAG 11772
841 snAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
   ::::: |||||
11771 CTAGGAATGAAAGAGAGAAACAAGAAATGCGAGAAATTAACACGCTGAA 11722
858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSer..... 871
   ::::: |||||
11721 GCGAGGAAA.....CTATTAGAAGAATCTAAAGAAAGCGCTTAAGGC 11681
872 .....LysProAspM 875
11680 TTATTTGGATTGCGTATCAAAAGCCAAAACGAGCTGAAGAAAGAAAT 11631
875 etSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGln 891
   ::::: |||||
11630 GCGAGAAATGCTCACCCCTGAAGCGAGAAAGCTATTAGAAGAGCTAAA 11581
892 GluProCysTyrHisGluIleIleThrLeuGluGlnThrGlnLeuCys... 907
   |||
11580 GAG.....AGTGTTAAGGCTTACAAAGACTCGCT 11552
908 .AlaLeuAlaIleAsnAsp.....GluCysTyrGlnValArgG 920

```

```

   ::  |||  ::::: |||||
11551 ATCAAGAGCTAGGAATGAAAGAGAGAAACAAGAAATGCGAGAAATTACTCA 11502
920 lnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
   |||  ::  ::::: |||||
11501 CGCTGAAGCGAAAAAACTA..... 11482
937 LeuGluTyrMetAlaIleCysAlaLeu..... 945
   |||||  ::  ::::: |||||
11481 TTAGAGAAATCAAGCGCTAGATTGTTTGAANAACGCTAAACCGAAGCTGA 11432
946 .....CysAlaLysAspProValLys.....GluArgArgAlaH 957
   |||  ::  ::::: |||||
11431 GAAAAAAGGTGCTCAAGATCTCCCTAAAGACTTACAGAAAAAGGTTT 11382
957 isAlaArgGln.....CysLeu.....ValLys 964
   |||||  ::  ::::: |||||
11381 TAGCAAAAGAGAGTGTTAGGTTTATTGTTGATTGCTATCAAAAGCCAAA 11332
965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerG1 981
   |||  ::  ::::: |||||
11331 AAGCAAGCTGAAAGAAAAAGAAATGCGAGAAATTGCTCACCCCTGAAGCGAG 11282
981 uLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisL 998
   ::::: |||||
11281 AAAGCTATTAGAAGCAAGCTAAAGAGAGTGTTAAGGCTTACAAA...GACT 11235
998 euLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeu 1014
   ::::: |||||
11234 CGGTATCAAGAGCTAGGAATGAAAGAGAGAAACAAGAATGCGAGAAATTA 11185
1015 .....LysAspValLysGluCysLe 1021
   ::::: |||||
11184 CTCACCCCTGAAGCTAGGAAACTATTAGACGCAAGAACTTAAAAAGACGCT 11135
1021 utPhePheValLeuGlu...IleLeuMetAlaLysAsnGluAsnAsnSerH 1037
   |||||  ::  ::::: |||||
11134 TAAGGCTTATTAGACTGCGTTTCAAGAGCTAGGAATGAAAGAGAGAAC 11085
1037 isAlaPhe.....IleArgLysMetValGluAsn 1046
   ::  ::  ::::: |||||
11084 AAGAATGCGAGAAATTAACACGCTGAAGCGAGGAAACTTTTAGAGAAAT 11035
1047 .....IleLysGlnThrLysAspAlaGlnGlyProAspAs 1058
11034 CAAGCGCTAGATTGTTGAAAGACGCTAAACCCGAACT..... 10996
1058 pAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnI 1075
   ::::: |||||
10995 .....GAGAAAAAAGGTGCTGTCAAAGATCTCCCTAAAGACT 10959
1075 lelleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspPro 1091
   ::  ::  ::::: |||||
10958 TGCAGAAAAAGGTTTACGTAAAGAGAGCGTTAAGGCTTATTAGACTGC 10909
1092 ValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnTh 1108
   |||  ::  ::::: |||||
10908 GTATCAAGAGCTAGG.....AATGAAAGAGAGAAAAAAGAAATG 10871
1108 rLysAsnTyrLeuProGluMetLysSerPheThrProGlyLysP 1125
   ::::: |||||
10870 CGAGAAATTCCTCACGCTGAAGCGAGAAACTATTAGAAGAACTATAAA 10821
1125 roLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAla 1141
   ::::: |||||
10820 AAAGCGTTAAGGCTTATTGTTGATTGCTATCAAAAGCCAAAAGCAAGCT 10771
1142 GlyLysGlnSerGlnThrLysSerSerArgMetGluThr.....Va 1155
   |||||  ::  ::::: |||||
10770 GAAAAAAGAAATGCGAGAAATTGCTCACACCTGAAGCGAGAAAGCTATT 10721
1155 lserAsnAlaSerSerSerAsnProSerSerProGlyArgIleLysG 1172
   ::::: |||||

```



```

13428 AATATGATCGAT..... 13417
93  oGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheI 110
   |||:|||||:|
13416 .....AAGAACTTTTCCTC... 13402
110  leThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
   :|||:|||||:|
13401 .....ANTGACGATCCCAATAGAACC 13381
127  ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnI 143
   ||| ||||| ||||| |||||
13380 TRATACAACTATTG.....AATATTGCA..... 13357
143  eCysPheGluLeuGluAspSerAsnGlnIlePheThrGlnLeuTyrArg 160
   |||:|||||:|
13356 .....GAAATTGAGCACAACCAACCATG.....AGAG 13329
160  hrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHis 176
   :|||:|||||:|
13328 CCTTTTATGAATGATAGTAATGTTGGCAACTATGAAGAATGTTTGAAG 13279
177  MetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerG 193
   :|||:|
13278 CTTATC..... 13273
193  nGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnL 210
   |||:|
13272 .....AAGACAAAAAAC 13260
210  euAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAla 226
   |||:|
13259 TTCAAGATCAG.....ATGAAAAAGACTCTA 13234
227  GlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMetLe 243
   :|||:|
13233 GAGGCT.....TATATGACTGCATCAAAA 13208
243  uGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleL 260
   :|||:|
13207 TGCCAAACTGAAGAAGAAAGGATCAAG.....TGTTTAGATTAAATCA 13164
260  euGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGln 276
   :|||:|
13163 AA.....GATGAAACCTGAAAAAAGCTTACTG..... 13135
277  LeuGluPheLysLeuLysSerAsnAspGluGluArgLeuGlnValVa 293
   |||:|||||:|
13134 .....ACCCACAAAAAGTTCAAGTGGC 13112
293  lLysLeuLeuAlaLysMetPheClyAlaLysAspSerGluLeuAlaSerG 310
   :|||:|
13111 T.....CTAGATTGTTTGAACAAACGCTAAACCCGATGAA..... 13078
310  lnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
13078 ..... 13078
327  ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAs 343
   ||| ||||| ||||| |||||
13077 .....GAACGGAAAGAGTGCCTAANA.....CTCATAAA 13049
343  nHisPro.....AspLeuAlaLysA 350
   ||| |||
13048 TGACCCCTGAGATTAGAGAAATTCGGTAAGGAATTAGAGCTTCAAAAAG 12999
350  spleuThrGluTyrLeuLysValArgSerHisaspProGluGluAlaIle 366
   :||| |||||
12998 AGCTTCAAGAGTAT..... 12985
367  ArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLe 383
   ||| |||
12984 ...AAGATTGTATCAAAACGCCCAACAGAGCTGAGAAAAAC..... 12943

```

```

383  uLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspL 400
12943 ..... 12943
400  ysArgTrpArgValArgLysGluAlaMetMetClyLeuAlaGlnIleTyr 416
   |||:|||||:|
12942 .....GAATGCTTGAAGGCTTGCT..... 12922
417  LysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnI 433
   |||:|||||:|
12921 .....AAGAAGCCATAGAAAGA... 12904
433  eAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleA 450
   :|||:|||||:|
12903 .....TTGAAACAACAAGCGCTAGATTGTTTGAACAAACGCTAAACCG 12861
450  sPaspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis 466
   |||:|
12860 ATGAA.....GAACGGAAGAGTGTATAAAATATTCCTCCAA 12823
467  AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaTh 483
   :|||:|
12822 GACTTGCAAAAAGAA.....CTACTAGCTGA 12797
483  rLeuAspLeuAsnAlaValLys.....AlaLeuAsnGluM 495
   :|||:|
12796 TATGAGCGTTAAGGCTTACAAAGACTGCGTATCAAGAGCTAGGAATGAA 12747
495  etrTrpLys.....CysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509
   ||| |||:|||||:|
12746 AAGAGAAACAAGAATCGAGAAATTAATCAACGCTGAAGCGAAAAAACA 12697
510 .....LeuAspLeuIleLysGlnProLysThrAspAlase 521
   ||| |||:|||||:|
12696 TTAGAGAANTCAAGCGCTAGATTGTTTGAACAAAGCTTAAACCGATCAAGA 12647
521  rValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProA 538
   :|||:|
12646 ACGGAAAGAGTGCTTG.....AAAAATCTCCCTA 12618
538  spProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlu 554
   :|||:|
12617 AAGACTTACAAAGCGATATTTTAGCTAAGAGAGAGCTGAAAGCTTATAA 12568
555  Asp.....AspGluLysIleArgLysGlnLeuGluVa 565
   ||| |||:|||||:|
12567 GACTGCGCATCTCAAGCCCAAACTGAAGCTGAAAAAAGAAATGCGAGAA 12518
565  lLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgG 582
   |||:|||||:|
12517 ATTACTCAGCGCT.....GAAGCGAAAAAATTTTAGAAG 12483
582  luIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeu 598
   |||:|
12482 AAGAAGCCAAA..... 12472
599  GluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspTh 615
   ||| |||:|||||:|
12471 GAGAGCGTTAAGGCTTATTGGATTGCGTATCTCAGGCCCAAACTGAAGC 12422
615  rGlu..... 616
   |||
12421 TGAACAAAAAAGAAATGCGAGAAATTTGCTCACACCTGAAGCGAAAAAAGT 12372
617 .....SerIleSerAlaLeuIleLysGlnValAsnLys 627
12371 TAGAAGAACCCAAAAGAGCGGTAGAGCTTATTGGATTGCGGTATCAAAA 12322
628  SerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspG 644
   :|||:|
12321 GCCAAAACGAAGCTCAAGAAAAAGAAATTCGAGAAATTAATCAACGCTGA 12272

```


10100 ACITTTTGAAG...GAAGCTAAAAAAGCGTTA 10072

671 luSerLeuLeuAlaCysLeu...LysMetAspAspGluLys... 683

10071 AGGCTTATTTGGATTGGTATCTCAAGCCAAAACCTCAAGATGAGAAAAA 10022

584ValAlaGluAl 587

10021 GAATGCGAGAAATTACTACCCCTGAAGCGAAAAAACTTTTAGAGCAACA 9972

687 aAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu..... 700

9971 AGCCCTAGATTGTTGAAAAACCTAAACCGAAGCTGATAAAAAAGGT 9922

701GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715

9921 GTGTCAAAAGATCTCCCT.....AAAGACTTACAGAAAAAGGTTTAGCT 9878

716 HisLysSerLysLysGlyProProArgGlnAlaLysTyAlaIleHisCy 732

9877 AAAAAAGAGC.....GTCAAGCGCTTATTTAGACTG 9849

732 sIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheG 749

9848 CGTAAAGAGCTAGGAATGAAAAAGAGAAAAA..... 9815

749 luProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThr 765

9814GAATGCGAGAAATTTACTCACC 9794

766 Pro.....LeuValThrIleGlyHisIleAlaLeuLeuAlaPr 778

9793 CCTGAAGCGAAAAAACTTTTAGAAGAAGCTAAAGAGAGCGCTGAAAGCT... 9746

778 oAspGlnPheAlaIleProTrpLysSerTrpValAlaThrPheIleValL 795

9745TATA 9742

795 ysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu 811

9741 AAGACTGCCCTCTCAAGCTAGAAATGAAGAAGAAAGGAGA..... 9701

812 TrpValProAspGluGluValSerProGluThrMetValLysIleGlnAl 828

9700 ...GCTTCGAGAAATTTACTCACCCCTGAAGCGAGAAACTATTAGAGCA 9654

828 aIleGlyMetMetValArgTrpLeuLeuGlyMetLysAsnHisSerL 845

9653 AGCTAAAAAAGCGTTAAAGCTTATTGGATTGCGTATCTCAA..... 9611

845 ysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAsp 861

9610GCCAAAAACGAA 9599

862 GlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLe 878

9598 GCTGAAGAAAAAGAACTCGGAAATTTACTCACCCCTGAAGCGAAAAACT 9549

878 uArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysT 895

9548 TTTAGAAGAAGCTAAAGAGAGCCTT.....AAAGCTTATAAAGACTGCC 9505

895 yrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIle 911

9504 TCTCTCAAGCTAGAAATGAACCTGAAGAGAGAGCTTCGCGAGAAATTTACTC 9455

912 AsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisL 928

9454 ACCCTCGAA.....GCGAGAAACTATTAGACACAGAAGTTAAAAA 9414

928 sGly.....LeuSerArgLeuArgLeuProLeuG 938

9413 GAGCGCTCAAGCGCTTATTTAGACTGCGTTTCAAGAGCTAGGAATGAAAAAG 9364

938 luTyrMetalaIleCysAlaLeuCysAlaLysAspProValLysGluArg 954

9363 AGAAACAAGAATCGAGAAATTTACTACGCTCAAGCAAGCAAGAAATTTT 9314

955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgG 971

9313 GAGAAGAAGCATTAAGTGTGTTGGAAAA.....GCTAGAATGA 9273

971 uTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuP 988

9272 A.....GAAGAAAGAAACATGTTTTTAAAAATCTTCTTAAAGATTTAC 9229

988 roGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004

9228 AGAAAAATGTTTTAGCTAAA..... 9209

1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLe 1021

9208GAGAGCTTAAAGCTTATAAAGACTGCCT 9180

1021 uTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHis 1038

9179 C.....TCTCAAGCTAGAAATGAAGAAGAAGAGAG 9148

1038 laPhe.....IleArgLysMetValGluAsnIle 1047

9147 CTTCGCGAAGAACTACTCACCCCTGAACCGAGAAACTCTTTAGAG..... 9104

1048 LysGlnThrLysAspAlaGlnGlyProAspAlaLysMetAsnGluL 1064

9103CAAGAAAGTTAAGAAGAGCGTTAA 9081

1064 sLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerT 1081

9080 GGCTTATTTGGATTGCGTTTCAAGAGCTAGGAATGAAAAAGAAAAACAGG 9031

1081 hrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhe 1097

9030 AATCGGAGAAATTTGCTCACCCCTGAACCGAGGAAACTTTTAGAGAAAGAA 8981

1098 PheThrGlnProAspLysAsnPheSerAsn...ThrLysAsnTyrLeuPr 1113

8980 CTCACAACAAAAAGATAAAGCGATAAAGATTTGCTTAAAAAACCCGATCC 8931

1113 O.....Prog 1115

8930 TAACACAGACGCGCTATCATGAAGTGTTTGGATGTTTGAGCGATCAAG 8881

1115 luMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeu 1131

8880 AAAAGCTTAATATTTCGAAGAGCTAGAGAAAAAGGCTGTTCGGATTGT 8831

1132 GlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLy 1148

8830 TTGCTATGCTATAAACCGATGAAGAAAAAGGAAATGCCAAACCTTTA 8781

1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProS 1165

8780 TAGTGATTGTATCCAGAAATTCCAAAATAAAGAGGCGACAGAACAAACA 8731

1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGlu...MetAsp 1180

8730 ATCAATTTAGTAAACAGAAAGGTTTGCATCAAGCAGCGAGTCTCTGGAT 8681

1181 HisSerGluAsn.....GluAspTyrThrMetSerSerProLeuProG 1195

8680 AACTTAGATGACCTACTGATCAACAGCCATAGAGCAATGTTTAGAGGG 8631

1195 yLysLysSerAspLysArg.....A 1202

8630 CTTGAGCGATGTAAAGGCGCTTAATTTCTAGGAATTTAAACGACAAAGCTG 8581


```

142 nLeuCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrA 159
11629 .....GAAATGAGCAAAACCATTTG.....A 11605
159 rgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMet 175
11604 GAGCCTTTATGAGTGTATGTTGATGATGGTCACACTATCAAGAAATGTTTG 11555
176 HisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSe 192
11554 AAGCTTATC..... 11546
192 rgLnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysA 209
11545 .....AAGACAAAA 11536
209 snLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThr 225
11535 AACTTCAAGATCAA.....ATGAAAAGACT 11510
226 AlaGlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMe 242
11509 CTAGAGGCT.....TATAACGACTGCATCAA 11484
242 tleuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuI 259
11483 AAATGCTAAACTGAAGAAGGATCAAG.....TGTTAGATTAA 11440
259 leuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
11439 TCAAA.....GATGAACCTGAAAAAGCTTACTGAAC 11405
276 Gln.....LeuGluPheLysLeuLysSerAsnAspAs 286
11404 CAACAAAAAGTTCAAGTGGCGCTAGATTGTTGAAAAACGCTAAAAACCGA 11355
286 nGluGluArgLeuGlnValValLysLeu..... 296
11354 TGAAGAACGAAACGAGTGTCTGAACCTCATNAATGACCCTGAGATTAGAG 11305
297 .....AlaLysMetPheGly..... 301
11304 AGAAATTCGTAAGGAATTAGGCTTCAAAAAGAGCTTCAAGAGTATAAG 11255
302 .....AlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpG1 316
11254 GATTGTATCAAAACGCCAAAACAGAGGCTGAGAAAAAT.....GA 11214
316 nCysTyrLeuGlyArgPheAsnAspIleHisValProIleArg..... 330
11213 ATGCTTGAAGGCTGTGCTAAAGAAGCTATAGAAAGATTGAAACAGCAAG 11164
331 ..LeuGluCysValLysPheAla.....SerHisCys 340
11163 CCCTAGATTGTTGAAAAACGCTAAAAACGATCAAGAACGAAACGAGTGC 11114
341 LeuMetAsnHisPro...AspLeuAlaLysAspLeuThrGluTyrLeuLys 356
11113 TTGAAAAATATTCCCAAGACTTGGCAAAAAGAACTACTAGCTGATATGAG 11064
356 sValArgSerHisAspPro.....GluGluAlaIleA 367
11063 CGTCAAGGCTTATAGGACTGCGTTTCAAAAGCTAGGAATGAAAGAGAGA 11014
367 rgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeu 383
11013 AAAAAGATGCCAGAAATGTCTACGCCCTGAAGCCGAGGAGAAAGTTA... 10967
384 LeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
10966 .....GAACAACAGGTTCTAGATTGTTTGA AAAACGCTAAAAACCGATGA 10923
400 sArgTrpArgValArgLysGluAlaMetMetGlyLeu..... 412
10922 AGAA.....CGAAAAAAGTGTGTAAGAGATCTCCCTAAAGACTTAC 10882
413 ..AlaGlnIleTyrLysLysTyrAlaLeuGlnSer..... 423
10881 AAGCGATATTTAGCTAAAGAGAGCCTGAAAGCTTATAAGATTCGCTA 10832
424 .....AlaAlaGlyLysAspAlaAlaLysGlnIleAl 434
10831 TCTCAAGCCAAAACCGAAGCTGAGAAAAAAGAAATGCGAGAAATTCGCTAC 10782
434 atrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp. 450
10781 CCTCAAGCGAAAAAAGCTTTTGAAGAAGAACCAAGAGAGCGTTAAAG 10732
451 .....AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetVal 464
10731 CTTATTTGGATTGCGTATCTCAAGCCAAAAGCTGAAGCTGAGAAAAAGAA 10682
465 ProHisAsnLeuGluThrThrGluArgMetLysCysLeu..... 477
10681 TCGGAGAAATTTACTCACCCCTGAAGCGAAAAAAGTTAGAAAGAGCTAA 10632
478 .TyrTyrLeuTyrAlaThrLeuAsp...LeuAsnAlaValLysAlaLeuA 493
10631 AAAAAAGCGTTAAGGCTTATTTGGATTGCGTATCTCAAGCCAAAAGCTGAAG 10582
493 snGluMetTrpLysCysGlnAsnLeuArgHisGlnValLysAspLeu 509
10581 ATGAAAAAAGAAATGCGAGAAATTTACTCACCCCTGAAGCGAAAAAAGCTT 10532
510 .....LeuAspLeuIleLysGlnProLysThrAspAlaSe 521
10531 TTAGACAAACAGCGCTAGATTGTTGAAAAACGCTAAAAACCGAAGCTGA 10482
521 rValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProA 538
10481 TAAAAAAGG.....TGTGTCAAGAGATCTCCCTA 10453
538 spProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlu 554
10452 AAGACTTGCAGAAAAAGTTTGTAGCCAAAGAGAGCGTTAGGCTTTATTWG 10403
555 Asp.....AspGluLysIleArgLysGlnLeuGluVa 565
10402 GATTGCGTATCAAGACGCCAAAAACGAAGCTGAAGAAAAAGAAATGCGAGAA 10353
565 lLeuValSerProThrCys.....SerCysLysGlnAlag 577
10352 ATTACTCACCCCTGAAGCAGAAAACTATTAGAAGAGCTAAAAAAGCG 10303
577 luGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnPro 593
10302 TTAAAGCTTATAAAGACTCGCTGTCAAGACCGCAGGAATCAAAAAGAGAAA 10253
594 ThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPr 610
10252 AAA.....GAATGCGAGAAATTTACTC..... 10232
610 ovalHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnL 627
10231 .....ACCCCTGAAGCAGAAAACTATTAGAAGAGCTAAAAA 10195
627 ysSerIleAspGlyThrAlaAsp.....AspGluAsp 637
10194 AAGCGTTAAACCTTATTTGGATTGCGTATCTCAAGCCAAAAGCTGAAGAT 10145
638 GluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLys 654
10144 GAGAAAAAAGAAATGCGAGAAATTTACTACCCCT.....GAAGCGAAAAA 10101
654 sValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheG 671

```



```
4318 AAACCCAAAGATTACGACTACTAAGGATAGAAATAGAAATGCTCGGATT 4367
1246 rGlnArgSerArgLysArgGlyHis.....ThrAlaSerG 1258
      ::::::::::::::::::::
4368 GTCCAAAGACTTCAGAAGATCATGATGAATGAATCTGTAGCTAAGG 4417
      ::::::::::::::::::::
1258 LuSerAspGluGlnGln.....TrpProGluGluLys 1268
      ::::::::::::::::::::
4418 AGAAGATGACCTACAGAGGCTCAAGAGATTCTCAATCTGAAAGTGAC 4467
      ::::::::::::::::::::
1269 ArgLeuLysGluAspLe.....LeuGluAsnG 1278
      ::::::::::::::::::::
4468 CAGCTCAAGAAACATAAAGAAATGTAGCTAAACACTGGAACACTGA 4517
      ::::::::::::::::::::
1278 uAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPro. 1294
      ::::::::::::::::::::
4518 AGAGGAACCTAAAGTTGCTCATTTGTCCTGAAGAACAGAGAACTA 4567
      ::::::::::::::::::::
1295 ..LysProLeuGlyGlyThrProLysGluGluProThrMetLysThr 1310
      ::::::::::::::::::::
4568 TTAATGAGTTAAGAGTGAATCTTTCAGAGAAGGAAACTGAAATATCAACC 4617
      ::::::::::::::::::::
1311 SerLysLys.....GlySerLysLysLysSerGlyProPr 1322
      ::::::::::::::::::::
4618 ATTCAAAAGCAGTTAGAACATCAATGATATAATACAGAACAGATCCA 4667
      ::::::::::::::::::::
1322 oAlaProGluGluGluGluGlu.....GluArgGlnSerGlyA 1336
      ::::::::::::::::::::
4668 AGAGATTATGAGAAGAGAGAACAACTTAATATAAACAAATGATGAGG 4717
      ::::::::::::::::::::
1336 snThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
      ::::::::::::::::::::
4718 TTCAGGAAACCTGATGAACCTGAACAACTTCAAGGAGCATCGCAAGCC 4767
      ::::::::::::::::::::
1353 GlnGlnArgAla...GluSerProGluSerSerAlaIleGluSerThr 1367
      ::::::::::::::::::::
4768 AGGATTCAGCTCACTCAAAAGTATAGAAAGTAAAGTCTCGAGTTGACC 4815
      ::::::::::::::::::::
seq_name: gb_ba2:AF282852

seq_documentation_block:
LOCUS AF282852 21347 bp DNA BCT 07-AUG-2000
DEFINITION Helicobacter pylori cagII pathogenicity island, complete sequence.
ACCESSION AF282852 U60176
VERSION AF282852.1 GI:8843969
KEYWORDS .
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori.
REFERENCE
  1 (bases 1 to 21347)
    Censini,S., Lange,C., Xiang,Z., Crabtree,J.E., Ghiara,P.,
    Borodovsky,M., Rappuoli,R. and Covacci,A.
    cag, a pathogenicity island of Helicobacter pylori, encodes type
    I-specific and disease-associated virulence factors
    Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14648-14653 (1996)
  2 (bases 1 to 21347)
    Covacci,A. and Rappuoli,R.
    Tyrosine-phosphorylated bacterial proteins: Trojan horses for the
    host cell
    J. Exp. Med. 191 (4), 587-592 (2000)
  3 (bases 17223 to 21347)
    Censini,S., Rappuoli,R., Lange,C. and Covacci,A.
    Direct Submission
    Submitted (06-JUN-1996) Molecular Biology, Chiron-Biocrine, Via
    Fiorentina 1, Siena, SI 53100, Italy
  4 (bases 1 to 21347)
    Censini,S., Rappuoli,R. and Covacci,A.
    Direct Submission
    Submitted (17-MAR-2000) Molecular Biology, Chiron-Biocrine, Via
    Fiorentina 1, Siena, SI 53100, Italy

REMARK
FEATURES
      Source
      1..21347
      /organism="Helicobacter pylori"
      /strain="CCUG 17874; NCTC 11638"
      /db_xref="taxon:210"
      /note="type I-specific and disease-associated virulence
      factor genes; partial sequence of cag pathogenicity island
      presented in AF282853"
      312..342
      /note="cag pathogenicity island integration signal"
      343..19467
      /note="cagII pathogenicity island"
      474..821
      /gene="cag-zeta"
      474..821
      /gene="cag-zeta"
      /note="HP520; cagI; JHP469; ORF6"
      /codon_start=1
      /transl_table=11
      /product="Cag-zeta"
      /protein_id="AAF80191.1"
      /db_xref="GI:8843970"
      /translation="MADTINTTETHETKPNFVNFKNLTDKRYDSLGLIGAGVL
      CCVLSGAMGIVGIIFVAIGIFLSPFNINLVKLEKSKKSKAAATVNNETQKSQATS
      VTNEPTAKETKD"
      929..1279
      /gene="cag-epsilon"
      929..1279
      /gene="cag-epsilon"
      /note="HP521; cag2; JHP470; ORF7"
      /codon_start=1
      /transl_table=11
      /product="Cag-epsilon"
      /protein_id="AAF80192.1"
      /db_xref="GI:8843971"
      /translation="MIQRELNSVFGCSNYVNRGCKGVLRREINGSKMYCLASVKA
      NIKENSPFLGCKNYLKCKWTASVDSQDPKYPKCNGLMKRKNKNEFFTATSLTKC
      NGILFLYQSEKKN"
      1279..2724
      /gene="cag-delta"
      1279..2724
      /gene="cag-delta"
      /note="HP522; cag3; JHP471; ORF8"
      /codon_start=1
      /transl_table=11
      /product="Cag-delta"
      /protein_id="AAF80193.1"
      /db_xref="GI:8843972"
      /translation="MFRKLATAVSLIGLLASNTLYAKEISEADKVIKATKTKETKKE
      AKRLKKAQRQIQPDHKPOYASVDDTKTQALFDIVDTLVNNDKSGDFGNSALDK
      KTYLYAMDLIDYNNVLSLENPDIKTTRAMGTADLIITGSLSEOVYNYNLLKLNKR
      AKFVLKINENMPYAQATFLRPKRSDNHAHTLDGAAIDENKLFQQRATFNANDV
      ICRPNDEVCSPLRDEMAMPTSDSKORKNIIAPYSYLRKETNNANQAQSPATQT
      APENSKELIEELIANSQLTANEEERKKLLAEKEQAEALAKYKLKLDENQKKLV
      EAEKKNNAKPRVPEVPSQTSNDETRVKEKENYNGLLVKKETTIRKSYEGTL
      ISENSKKTPLNPNDLRLSEETIKSVYIKSNGLCYTNGINLYVKIKNDPVKEGMLCG
      YESVQNLSPDKDKLYDKQKLQKALLKDSK"
      2729..3238
      /gene="cag-gamma"
      2729..3238
      /gene="cag-gamma"
      /note="HP523; cag4; JHP472; ORF9"
      /codon_start=1
      /transl_table=11
      /product="Cag-gamma"
      /protein_id="AAF80194.1"
      /db_xref="GI:8843973"
      /translation="MFGKWLTGLTFLNSLAYPCOKVITTFKQYENLTQIROKGDNEV
      VCPPLISLTALLESLGLNNKREKSLKDTYSMFHITNTAKKFYPTYSKTLLYKLLN
      DVDFAIQLAKILKENFDYKQKYPNKSIVQLVEMAIGTYNGMKHNPNGAYVKKFR
      IYSQVRYNE"
```


3541 AGTGAGATGCCAGAAAAGATTAAATCAATAGAGAAATTTAAAGAAATGAATT 3590
 982 LysLeuLeuSerLeuLeuProGluTyrValValProTyThrIleHisL 998
 3591 AAAGAACAAAGAAATGACATTGGACACATATGGAACACAGAGCGCTTGAGT 3640
 998 euLeuAlaHis... AspProAspTyrValLysValGlnAspIle... Glu 1012
 3641 TGGCTCAGAAACTTAATCAAAATATGAGGAAGTGAATCTATAACCCAAA 3690
 1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluLeuLeuMe 1029
 3691 GAAAGAAAAGTTCTAAAGGAA... ..TTACAGAGATCAT 3725
 1029 talLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluA 1046
 3726 TGAACACAGAGAGAGACCCTTAGAGGATATTAAGAGAATTTGAAGCTA 3775
 1046 snIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn 1062
 3776 CAGGCCTACAAACCAAGAACACTAAAAATTGCTCATATTCACCTAAAA 3825
 1063 GlnLysLeuTyThrValCysAspValAlaMetAsnIleLeuMetSerLy 1079
 3826 GAACACCAAGAAACTATTGATGACTAAGAAGAAGCGTA... .TCTGA 3869
 1079 sSerThrThrTySerLeuGluSerProLysAspProValLeuProAlaA 1096
 3870 GAAGACA... ..GCTC 3880
 1096 rgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyLeu 1112
 3881 AAATAATAAATCTACGAGCATTAGAAAANTCCCATACCAA... .TTA 3924
 1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAs 1129
 3925 CAAGAAGAGATCCCA... .. 3939
 1129 nValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerG 1146
 3940 .GTGCTTCATGAGGAAGATTACTGCCCTAAATGAAAAGAGTCAGT 3988
 1146 InThrLysSer... .SerArgMetGluThrValSerAsnAlaSerSer 1160
 3989 AGACTCAGGAACACATGATGAATGAGTGGATTATTAACAGAACACTCCACA 4038
 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSe 1177
 4039 ACCAAGGACTCAACAACACTGCGCAAGATA... .. 4068
 1177 rGluMetAspHisSerGlu... .AsnGluAspTyThrMetSerSerProL 1193
 4069 .GAAATGGAAGGCTCAGGTTCAATGAAAAATTTCAAGAAAGTCAGGAAG 4117
 1193 euProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
 4118 AGATAAATCTCTAACCAANGAAGAGACACACCTTAAACGATAAAGAA 4167
 1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu... 1224
 4168 GCCCTTGAAGTTTAAACATGACCAGCTGAAAGAACATATTTAGAGAAACTTT 4217
 1225GlnGluGluLysLeuGlyMet... 1231
 4218 GGCTAAAAATCCAGGAGTCTCAAGCAAAACAGACAGACTCCTTAAATATGA 4267
 1232AspAspLeuThrLysLeuValGlnGluGln... 1241
 4268 AGAAAAAGACAACTGCAAAATTCGTGATGAGATGGACCAATTC 4317
 1242 LysProLys... ..Glyse 1246

1376 ACAAATGAGAACTCAAC..... 1395
 153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHi 169
 1396TATCGAGATCAATTAATATACCAACAAAT...AT 1427
 169 sAsnGlnLysValHisMetHisMetValAspLeuMetSerSeriLe... 184
 1428 AACACAAAACACATAAGCTTTCTATAAATTTATTACGAGAAATTGATG 1477
 185IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
 1478 AATCTGCTGCTCGAGCTCGATGTTTTCAGTAACACTCTGTGATACATTA 1527
 200 Leu...ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaty 215
 1528 AGTGAGATAGATGAATCCAGCAACAAAGCTACTAAATCAGGAGAAATAT 1577
 215 rasPLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProT 232
 1578 AGAA..... 1581
 232 yrIleThrThrPheAsnGlnValLeuMetLeuGlyLysThrSeriLe 248
 1582AGTGAGTTG 1590
 249 SerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIl 265
 1591 AACTCACTCGCTGCTGACTATGATATCTGGTATTA..... 1626
 265 eAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuL 282
 1627 .GACTATGAACAACTACGACACAGAAAAAGAAATGGAATTGAAATTA 1675
 282 ySerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
 1676 AAGAAAGATGATTTGGATGAATTTGAGCTCTAGAAAGAAAACTAA 1725
 299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTr 315
 1726AAGATCAAGAGATGCAA..... 1743
 315 pGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuG 332
 1743 1743
 332 lucysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAla 348
 1744CTAATTCATGAATTTGCACTTA 1767
 349 LysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAl 365
 1768 AAGAATTTAGTTAAGCATCGAGATATATATCAAGATCTTGAGAGATGA 1817
 365 aIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAspI 382
 1818 ACTC.....AGTTCAAAAGTAGAGC 1837
 382 leLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
 1838 TGCTTAGAANAAGGAGACCATTAAGAGAGCTACAGGAATACATAGAC 1887
 399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIl 415
 1888 TCtCAAAAG.....CTAGAAAAATATAAAATGGACTTGtCA..... 1923
 415 eTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysG 432
 1924TACTCATGTGAAGACATTTGAAGACCCCAACAAATGAAGC 1963
 432 InIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSer 448
 1964 AGACTCTGTTGATGCTGAACACTGA..... 1989

449 IleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
 1990GCCCTTGATGCAAGAGAGAAATCAGCCTTTCTTAGAAG 2027
 465 oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrA 482
 2028 TGAATAATCTGGAGTTGAAGGAAAAAATGAAGAACTTGCACACTACATACA 2077
 482 leThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
 2078 AGCAAAATGGAA.....AATGATATTTCAGTTATAT 2106
 499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysG 515
 2107 CAAAGCAATTTGGAGCAAAAAAAGAAATGCAAGTTGATCTCGAGAAAGA 2156
 515 nProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValI 532
 2157 A.....TTACAATCTGCTTTTAATGAGATAACAAAC 2188
 532 leThrArg.....AsnLeu 536
 2189 TCACCTCCCTTATAGATGCAAGATTCCAAAGATTGCTCTGTAATTTC 2238
 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLe 553
 2239 GAATTGCAAGGAAGATTTACTGATCTTCAGAAAGAACTAAATAAGAAAGT 2288
 553 uGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProT 570
 2289 TGAAGAAATGAAGCTTTCGGGGAAGAAGTC...ATTTCGCTTTCA... 2331
 570 hrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
 2332GAATTGAATCTTTACCTTCTCAAGTAGAAGGCTGAGGAAAGAG 2376
 587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPh 603
 2377 ATA.....CAAGACAAATCTGAAGAGCTCCATATAATAACATC 2414
 603 eLeuLeuGluArgIleAlaProValHisIleAspThrGluSer...IleS 619
 2415 AGAAAGATATAATTTCTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTC 2464
 619 exAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp 635
 2465 AAGGTTTACTTGAAGAAATTTGGAAAAACAAAGATGACCTAGCAACTACA 2514
 636 GluAspGluGlyValProThrAspGlnAlaIle..... 646
 2515 CAGTCGAATTATAAAAGCACCTGATCAAGAATTTCCAAAATTTCAAACCCCT 2564
 647A 647
 2565 TCATATGGACTTTTGACCAAAAGTATAGATGGTCTTGGAGGAGATGAGA 2614
 647 rgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
 2615 GAATGAATCAGGAATAGTTAATCTCTCTAAA..... 2646
 664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAs 680
 2647GAAGCCCAAAATTTGATTCAGTTCAGTTGGTGGTTCGAGGACCGA 2690
 680 p.....AspGluLysValAlaGluAlaAlaL 689
 2691 GCTTTCTTACAGACCCCAAGAACTTCAGGAGAAACACGTCGAGGTCAAG 2740
 689 euGlnIlePheLysAsnThrGlySerLysIleGluAspPheProHis 705
 2741 AAAGACTA.....AATGAGATGGAACAGCTGAAGCAACAATTAGAAAT 2784

11643TCGACAAGAAGTAAGTTTATT 11663
794 ValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLy 810
||||| ||| |||
11664 GTTAAA.....CTTACTAAC..... 11678
810 LeuTrpValProAsp..... 815
||||| |||
11679TTTCATTTTGTGACTATATATGATGAATGATGACCAATCTTACGATTTG 11724
815 815
11725 TATTTCACAGGTTTATGAAGTGTCAATAAATCTTATGTAGAGTTAA 11774
816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMe 832
||||| ||| |||
11775 GAAGATGCATCAAGTCAGTGAATCTCTATTTGGTATTTTAAATATCTC 11824
832 tValArgTrpLeuLeuGlyMetLysAsnHisSerLysSerGlyThrS 849
||||| ||| |||
11825 TATTATTCATTAACGGTTTATATAGAAATTTTCCCA..... 11861
849 erThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThr 865
||||| ||| |||
11862CTGATGGCTCGCAGGTACATATCTCATGCCGATGT... 11900
866 GluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaI 882
||||| ||| |||
11901 GAAGAGCTTAATCTGAAGAAAGAACGCTTTTAAAGATAATCGAGGATAAATC 11950
882 aGlySerAlaIleValLysLeuGlnGluProCysTyrHisGluIleI 899
||||| ||| |||
11951 TTCAGCCAGTCAGGTGAATATTTCTTA.CATTCTATCATTCACCATCTT 11999
899 leThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys 915
||||| ||| |||
12000 TATATTTACCAAAATTTCAA...TGATCTGGTTTCCCTAAATAAATCTA 12046
916 TyrGlnValArgGlnValPheAlaGlnLysLeu..... 926
||||| ||| |||
12047 AGCAGGATAAGGAAGATGATCTGCTTGAGTCTACTCTTTATTCGCTTG 12096
927HisLysGlyLeuSerArgLeuArgLeu..... 935
||||| ||| |||
12097 TAGTGAAATCTCCATAGTCTATGATATACCCCAATTTATATTTTAAT 12146
936ProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 949
||||| ||| |||
12147 TTAGCTTTACTCTTGAGTTTGTGCTGT.....ATGTCAGT..... 12184
950 ProValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
||||| ||| |||
12185 ...ATACAAAGGAGAAATCCAAAGAGAGAAATTTGCTCTAAGAATGT 12231
966 eAsnValArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysL 983
||||| ||| |||
12232 GGAACCG..... 12238
983 euLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeu 999
||||| ||| |||
12239AGTAGTCTCCAGAA..... 12253
1000 AlaHisAspProAspTyrValLysValGlnAspIle..... 1011
||||| ||| |||
12254GTCAGGTATGAAGTATATATAAGAATCTAGTT 12285
1012GluGlnLeu.....LysAspValLysGluCysLeuTrpP 1023
||||| ||| |||
12286 TTAGTTGTTTGAAGTTTGTCCGTGAGTGAATTAGTTCCAAATATGTG. 12334
1023 heValLeuIleLeuMetAlaLysAsnGluAsnHisAlaPhe 1039
||||| ||| |||
12335ATGTAGATCTCTATGCAACAAATGAAGAAGAA..... 12367

1040 IleArg.LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyP 1056
||||| ||| |||
12368 ...AGACTCTGTAAACAGACTCCATTAAAGCAACAAACAAAGAACCAAGGTG 12414
1056 roAspAspAlaLysMetAsnGluLysLeuTyrThrValCysaspValAla 1072
||||| ||| |||
12415 CACTGAAGGCTCTAAGCAATGAACCA..... 12440
1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLy 1089
||||| ||| |||
12441GAAAGCACTACAGGGAATAATCTTAATCCTTGAA 12475
1089 sAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPhes 1106
||||| ||| |||
12476 A.....AAGCTGAATGCTGAACCTGATAAAACAAGAG 12507
1106 erAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrPro 1122
||||| ||| |||
12508 GCAGAACT..... 12515
1123 GlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSe 1139
||||| ||| |||
12516 GCAGAAACGAG..... 12527
1139 rSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValS 1156
||||| ||| |||
12528AAGGTGACTCAAGCTATGCACCGGAAATTCGAAAAA.... 12563
1156 erAsnAlaSerSerSerAsnProSerSerProGlyArgIleLysGly 1172
||||| ||| |||
12563 12563
1173 ArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMe 1189
||||| ||| |||
12564GATTGTGATGAGCAGGAGACCTCGAAACCAAGATCAACAAGA 12607
1189 tSerSerProLeuProGlyLysLysSerAspLysArgAspSerAspL 1206
||||| ||| |||
12608 CAGTCTGAAATTG...GGGAAAGAATCAGATGCA.....GAGCCTGATC 12648
1206 euValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
||||| ||| |||
12649 GTATGGAAGATCACCAGAAATTCCTGAAATCCAAATGTAGAAACCAA 12698
1223 ThrGluGlnGluLysLeuGlyMetAspLeuThrLysLeuValG1 1239
||||| ||| |||
12699 ACTGATGGAGAAGAG.....CA 12715
1239 nGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThra 1256
||||| ||| |||
12716 GGAGGCGAGCAAGAGCCCAACGCGCAGAGTCTAAACTAATGCGAGAGGC 12765
1256 lAserGluSerAspGluGlnGlnTrpProGluLysArgLeuLysGlu 1272
||||| ||| |||
12766 CAATTCGACAAACCCGAACTGATGGAAGAGCATATAATCATTTGAAGGAG 12815
1273 AspIleLeuGlu.....AsnGluAspGluGlnAsnSerProProLysLy 1287
||||| ||| |||
12816 CCAATGCGAGACCCCAATCTGTGTGAGAAGAGCAGGAGGCGACCAAGA 12865
1287 s.....GlyLysArgGlyArgProProLysP 1296
||||| ||| |||
12866 GCCAAATGCTGAGTCAAAACTGATGGAGAAAATCAGGAGGCGACCAAGA 12915
1296 roLeuGlyGlyThrProLysGluGluProThrMetLysThrSer... 1311
||||| ||| |||
12916 AGCTAACTCGACAACCGCAAACTGATGAGGAAGAGACCAAGGTAGTGAT 12965
1312LysLysGlySerLysLysSerGlyProProAlaProG1 1325
||||| ||| |||
12966 GAGGTAGACAAATGTCACAGAAAGAGCAATGTAGAACCCGGAACCTGA 13015


```

gene      14106..17166
CDS       join(14106..14238,14360..14514,15043..15228,15405..15757,
16959..17166)
/ gene="T24D18.5"
/ note="Strong similarity to cinnamoyl CoA reductase
gii12960364 from Populus balsamifera. ESTs gbIN95902,
gbIA1992693, gbIA1995837 come from this gene."
/ codon_start=-1
/ evidence=not experimental
/ protein_id="AAFI18492.1"
/ db_xref="GI:6587801"
/ translation="MEVDVASPAKTVCVTCAGGYIASWLVKILLERGVTVKGTVRNP
DDPNKTHLRELEGGKRLICKADLDQYEAALIDCGDGVFHTASVTDPPQWVEP
AVNGAKFVINAAEAQKRVVITSSIGAVYMDPNRDPEAVVDESCWDLDFCGSKNTKM
YCYGKMWAEQAWEVTAKEKGVLDVLPVLPLOPTINASLVHVKLYLTKPLTK
ANLTOAVDVDRVALAHVLYEAPSAQSVLLAESAHRGEVVEILAKLPEYLPPTK
CKDEKNPRAKYFTNQIKIDLGLEFTSTQSLYDVKSLQEKGHILAPPPPSASQE
SVENGKIGS"
/ complement(17453..20317)
/ gene="T24D18.6"
CDS       complement(join(17453..17770,17863..17958,18041..18172,
18248..18352,18449..18622,18698..18767,18839..19020,
19098..19207,19342..19431,19565..19643,19731..19771,
19858..19936,20225..20317))
/ note="Strong similarity to gi|2130080 Nrampl protein from
Oryza sativa and contains a Pf01566 Nrampl domain."
/ codon_start=1
/ evidence=not experimental
/ protein_id="AAFI18493.1"
/ db_xref="GI:6587802"
/ translation="MAAETASGNSRISNSPLIENSNOILVPEKKSKNFFSYLGP
GFLVSTAYIDGPNFTDQSGAQYKYLWIIIVASCAALVIOSLANLGVSKHILA
EHCRAEYKVPNFWVAVIYVAGLVGTAFALNMFNIPVWIGLVLTGLSTLILAL
QOYKIRLEFLIAFLVETALCFEVLHYKPKQEVLYGLFVPLQKNGATGLAISL
LGMVPHNLFHSALVLSRKIPRSVTGIREACRYLIESGLAMWAFNLVNSVYS
GAVCNASDLPEDRACQDLDLNAKASFLRNVVKWSSKLFALIALASGOSSTITGY
AGYVWOGFDLRLLEPLRFLRCLAIIPSLIVALIGGSAGAKLIIIASMLISPEL
PFALVPLKFTDSKTKMGSHANLIVSSVTWIIIGGLIMGINIYLVSSFKLLHSHM
NLVAIVLGLVGFSGIATYLAALSYLVLRKNRESSSTHFLDFNSQSTETLPRDIAN
MOLPNRVAVIGDLN"
/ complement(21795..23745)
/ gene="T24D18.7"
CDS       complement(join(21795..22106,22210..22351,22466..22549,
22637..22694,23283..23745))
/ note="Contains similarity to gi|112785
DNA-3-methyladenine glycosidase I from Escherichia coli."
/ codon_start=-1
/ evidence=not experimental
/ protein_id="AAFI18494.1"
/ db_xref="GI:6587803"
/ translation="MSVPPRRFVNSDEREFSRVLGPTGNKLQRPKPMKLEKPMMEK
TIIDSKDEKAKPTTPASPTTLKQCSLSSILRNKNSASMTASYSDSASSCESPL
SVASSCKKVPVRSQSVSTRKLSVGKEEKVSGDCFADGRKCAWITPKADPCIVA
FHEDEMGVPHDDKKLFLKLSGALAESWTDILSRHILREVFDFDVPVAELND
KKLTAPGTAAISLSEVKIRSIDLNSRHVKIIAECGSLKKYMNFMVNNKPTQSFYR
QROVPVTKRAEFISKDLVRRGRFSVPTVIYSFMQAAGLNDHLICFRYQCCVDA
ETTTTKAKKKNERESDK"
/ complement(24565..24989,25088..25238,25326..25516,25589..25846,
25955..26450)
/ gene="T24D18.8"
CDS       join(24565..24989,25088..25238,25326..25516,25589..25846,
25955..26450)
alignment_scores:
Quality: 230.50      Length: 1581
Ratio: 0.335        Gaps: 81
Percent Similarity: 43.580      Percent Identity: 19.418
alignment_block:
US-09-512-581-2 x T24D18 ..
Align seg 1/1 to: T24D18 from: 1 to: 80442

```

```

43 MetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLe 59
.....|.....|.....|.....|.....|.....|.....|.....|
9370 GTCGAGCAAGATCTTCTTCATCAGTACAAAAGGCCCTACATCCACCAAT 9419
.....|.....|.....|.....|.....|.....|.....|.....|
59 uAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspV 76
.....|.....|.....|.....|.....|.....|.....|.....|
9420 GAGGCTTTGGTATCCGCTGAT...CTTTTGAGAAACCCCTGATTCGTATG 9466
.....|.....|.....|.....|.....|.....|.....|.....|
76 aArgLeuLeuValAlaCysCysLeuAlaAspPheArgIleTyrAla 92
.....|.....|.....|.....|.....|.....|.....|.....|
9467 TTAGGTTTCTGTTCTCTCTGCTTGCACCGAAATATAGAGGATTTACTGCT 9516
.....|.....|.....|.....|.....|.....|.....|.....|
93 ProGluAlaProTyr...ThrSer..... 99
.....|.....|.....|.....|.....|.....|.....|.....|
9517 CCTGAGGCCCTTACAAACGATGAGCAAAATGAAGGTTGATATATCCATTAA 9566
.....|.....|.....|.....|.....|.....|.....|.....|
100 .....ProaspLys..... 102
.....|.....|.....|.....|.....|.....|.....|.....|
9567 CCACAATTGTGTACCAAGTTAAACGGTCTCTTGTGTTTGTCAATGCTTTC 9616
.....|.....|.....|.....|.....|.....|.....|.....|
103 .....LeuLysAs 105
.....|.....|.....|.....|.....|.....|.....|.....|
9617 TTCAATATTCTCTCATGAGGCATTGGCTTTCTTATTGCGCTTACAGGA 9666
.....|.....|.....|.....|.....|.....|.....|.....|
105 pIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysS 122
.....|.....|.....|.....|.....|.....|.....|.....|
9667 CATCTTCCAGGTGACAATAGAAAGCCTTTGAAAAACTTGCTGATGCTTCTA 9716
.....|.....|.....|.....|.....|.....|.....|.....|
122 erProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVal 138
.....|.....|.....|.....|.....|.....|.....|.....|
9717 GTCCGAGTTACAGGAAGCCGAGGTTATTCTTTGAGACTGTTGCAAGGTT 9766
.....|.....|.....|.....|.....|.....|.....|.....|
139 LysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIle..... 153
.....|.....|.....|.....|.....|.....|.....|.....|
9767 AGATCATCTTGGTTATGTTGAGCTTGGAGTCGCGAGGATCTTGTCTCTAGA 9816
.....|.....|.....|.....|.....|.....|.....|.....|
154 .....PheThrGlnLeuTyrArgThrLeuPheSerValIle. 165
.....|.....|.....|.....|.....|.....|.....|.....|
9817 AATGTTTCAACGGTCTTCAAAATCATAGGTAGATACAGAACCTGATT 9866
.....|.....|.....|.....|.....|.....|.....|.....|
166 AsnAsn..... 167
.....|.....|.....|.....|.....|.....|.....|.....|
9867 AACAATGCTCAGTTAACTTTAGGTACTTTATGTGACACTGATGATGGCT 9916
.....|.....|.....|.....|.....|.....|.....|.....|
168 .....Gly...HisAsnGlnLysValHisMet 175
.....|.....|.....|.....|.....|.....|.....|.....|
9917 TTTCTTTTGGGCATTTATCTAGCGCGGATCATCTCAACTGCTGCTTGTGC 9966
.....|.....|.....|.....|.....|.....|.....|.....|
176 HisMetValAspLeuMetSerSerIlelleCysGluGlyAspThrValSe 192
.....|.....|.....|.....|.....|.....|.....|.....|
9967 TCAATGGAAACGATTATGATCATCAGTTATAGATGAAAGTGAAGAAGTACC 10016
.....|.....|.....|.....|.....|.....|.....|.....|
192 rcGlnGluLeuLeuAspThrValLeuValAsn..... 202
.....|.....|.....|.....|.....|.....|.....|.....|
10017 CATGACTGCTCGAGATTCTCTTGACTACTCTCAAAAAGGACAGTCAGG 10066
.....|.....|.....|.....|.....|.....|.....|.....|
203 .....LeuValProAlaHisLys..... 208
.....|.....|.....|.....|.....|.....|.....|.....|
10067 TAAAGGCTTTTGTCTATTTTTCAGGTTAAATTTCTTCCAGCTTTTCAATTGACT 10116
.....|.....|.....|.....|.....|.....|.....|.....|
209 .....AsnLeuAsnLysGlnAl 214
.....|.....|.....|.....|.....|.....|.....|.....|
10117 TGAGCTTTTGATTTTCTGTCGGCATGCTTATAGGATGTCTCACCAGCTGC 10166
.....|.....|.....|.....|.....|.....|.....|.....|
214 atyAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluP 231
.....|.....|.....|.....|.....|.....|.....|.....|
10167 TTTGACGCTTGGGAGAGGTTCTCAGTAGTTCACCTGTAAGCTTCAGC 10216
.....|.....|.....|.....|.....|.....|.....|.....|
231 roTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSer 247
.....|.....|.....|.....|.....|.....|.....|.....|
10217 CGTGCAATG.....GAAGCTTTGAGCTCTCAGGACCAGC 10254
.....|.....|.....|.....|.....|.....|.....|.....|

```


5342 GCAGGAGCTGCAGAGCCACATTAAACAGCAGGACGACATCTAAGCAAGGAA 5391

1346 HisArgValSerArgAlaGlnGlnArgAlaGluSerProGlu 1360

5392 AAT.....GAGATTAAAGCGAGAATCTGAGCAA 5421

seq_name: gb_p14:T24D18

```

seq_documentation_block:
LOCUS      T24D18      80442 bp      DNA      PLN      16-DEC-1999
DEFINITION Arabidopsis thaliana chromosome 1 BAC T24D18 sequence, complete
sequence.
ACCESSION  AC010924
VERSION    AC010924.2  GI:6272370
KEYWORDS   HTG
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 80442)
AUTHORS   Liu,S.X., Yu,G., Sakano,H., Jhaveri,A., Lee,J.M., Lenz,C., Pham,P.,
            Toriumi,M., Chin,C., Chlou,J., Choi,E., Chung,M., Gonzalez,A.,
            Howng,B., Koo,T., Li,J., Liu,A., Vaysberg,M., Altafi,H., Brooks,S.,
            Buehler,E., Chao,Q., Conn,L., Conway,A., Hansen,N.,
            Johnson-Hopson,C., Khan,S., Kim,C., Lam,B., Nguyen,M., Palm,C.,
            Shinn,P., Tambunga,G., Davis,R.W., Ecker,J.R., Federspiel,N.A. and
            Theologis,A.
            The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1
            Unpublished
            2 (bases 1 to 80442)
            Theologis,A.
            Direct Submission
            Submitted (28-SEP-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
            3 (bases 1 to 80442)
            Theologis,A.
            Direct Submission
            Submitted (06-NOV-1999) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
            4 (bases 1 to 80442)
            Theologis,A.
            Direct Submission
            Submitted (16-DEC-1999) Plant Gene Expression Center, 800 Buchanan
            St., Albany, CA 94710, USA
            On Nov 6, 1999 this sequence version replaced gi:5931388.
            The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.
            The sequence does not represent the sequence of the entire insert
            of this clone. It is shorter by 13381 bp because we submit only
            the unique sequence of the clone. However, in order to facilitate
            the jointing of overlapping clones in the future for creation of
            larger contigs, we provide small overlaps (200 bp) between
            overlapping submitted clones. The 3' end of this sequence overlaps
            by 200 bp the 5' end of the sequence of the clone F309.
FEATURES             Location/Qualifiers
     1..80442
     1..80442
        /organism="Arabidopsis thaliana"
        /cultivar="Columbia"
        /db_xref="taxon:3702"
        /chromosome="1"
        /clone="T24D18"
        /cdate="1202..3974"
        /gene="T24D18.1"
        /join(1202..2102,2184..2365,2479..2568,2682..2969,
        3324..3536,3744..3974)
        /gene="T24D18.1"
        /note="Contains similarity to gb|AF136530 transcriptional
        regulator from Zea mays. ESTs gb|F14071, gb|Z26823,
        gb|A1998935 come from this gene."
        /codon_start=1
        /evidence=not_experimental
        /protein_id="AAF18488.1"
        /db_xref="GI:6587797"

```

```

/translation="MSISDEBAEISESEIEDYSETPYRLRLDRGTYKVKVNGQLRCPFC
AGKKQDYKELYAHATGVSQATRSALOKANHLAMFLENLACVAYEPVPPPV
VPQDDETPNPNHVVYPMGIVYVPLKEADDEKLLDLSAYLQTLTKFKPIEVNAF
WVEQDSIVGVYAKFNGMDSGAGATELEKEFETOGSSKKWERTSGDSKRAIGWCAR
ADDFESQPIGEIYLSKQGLRTVSDISQKNVDNRNTVLEELSDMIAMTNDLNRKQVYS
YNTAMSLQRIYDEKKNLHQAFADSTKKMQMSLRHIOKILYDKELSNELDKRNRDL
ESRAKQLEHEALTDLDRQKLEDEKRSKQMSLQSLASRQKKADESVLRKLEVEHQOR
OKDALNKILLKOLDPTQKLEMEIQELKGLKQVMKHLGDDDDDAVQKKKMEDEL
DDKRAELGLESMNSVLMTKERQSNDEIQAAKKLIAGLTGLLGAEITDGYKRWGELD
EKPFLLCKRYSANEAAREATIGCTWQENLKNFSWQPFKHEGRTGDAEEVVDDEDE
OLKLLKRWGKREYHNAVNTALVEMNEYNASGRYTPPELWNFKRGRKATLKKEVITFIS
DITLKKRRT"
5213..6073
/gene="T24D18.2"
5213..6073
/gene="T24D18.2"
/note="Similar to gi|Q60809 CCR4-associated factor 1
(CAF1) from Mus musculus. EST gb|Z26822 comes from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF18489.1"
/db_xref="GI:6587798"
/translation="MSQAPNPEEDDTIEIRVNVNHNLEQEMALIEQSIDDFPYVMD
TEFGIVCKTVTANPNPNPYSIHYEYNTLTKANVMKLQGLTLSDKGNLPTCG
TNKOCIQWFNFEENVISDMFALDSIELLRKSAIDLEKNNEGVDKAFKELMGSGV
VLYDKIHWTFCHGVDYFGYLLKLLSGKELPEISDFQMEKFFPVVDYIKYLMGFCF
NLNGGLEKIAELLGKVRGVGISHQAGSDSLTLRTFIKMKEFFFTGSLLLKSGFLGLD
NPRLITGSKN"
6953..7992
/gene="T24D18.3"
join(6953..6960,7349..7649,7771..7836,7933..7992)
/gene="T24D18.3"
/note="Strong similarity to gb|AF067732 ribosomal protein
S12 from Hordeum vulgare. ESTs gb|T41772, gb|T42570,
gb|A1999345, gb|T20784, gb|F20068 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF18490.1"
/db_xref="GI:6587799"
/translation="MSGDEAAFPVVPVPAEPAIPEMDMLTALTLRKRARAYGVY
VRGLHECAKLIKRVLAQVLAECNQPDYVKVKALCADHEVRLTVPKARTLGEWA
GLCKRDSGNARKVVGCSCLVVKDFGEETALSIVNKHIAQ"
8923..13301
/gene="T24D18.4"
join(8923..9036,9346..9549,9665..9846,9939..10065,
10150..10347,10469..10519,10604..11311,11398..11493,
11579..11652,11736..11787,11877..11963,12052..12095,
12184..12258,12341..13301)
/gene="T24D18.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF18491.1"
/db_xref="GI:6587800"
/translation="MGPIVEATEESKEQALDAAENLLKPHESTDATLSLLEVMESLL
ATVQDLSVVSKALHPMRAVLSADLNPDSVRSVSVCLTEIMRITAEAPYND
EQMKDIFOVTIEAFELKADSSRYRKAIEVLETVAKVRSSLVMDLDCDDLVLEMFQ
RFLKIRDPQLVLSVLEMTIMTVIDESEVPMDLLLEITLVTKKSDQSVSPAALTL
VEKVLSCCTCLQPCIMEALKSGSTSLDMYSPVSSICQSEFATQAHNDVYKPDNEA
DEKISEGOVVPDSLEKLNGLSRKGRSRKRGTRRANGDEKVTANGLSEST
DAETASGSTRKGMKPKSLMPEEGYSFKTSSKKVQEKELGDSLVAAKKVPLPS
KVGQTNQSVISLSSSGRGRSRKRTKMEETDHDVSVATQPAKQTKVKTNPAP
EDLTNRNVKHHEDGIGTKSKKRAKADNGLSAKPLAETMMVMPKSGKLVHSDAK
KNEGASMDTPIPOSSSKKKDSRATTPATKSEQAPKSPHKMRKIAAGEEVESITNE
LGEELVGRVNVWPLDRKFYGVKSYCRVKMHQVTSYSDGVSEELMLKREFKRIIE
KDSASDEKDDLLSTPLSAFIOREKSKKRVISKNVSPSSPVSRSSMOTMKKDS
VDSIKQTKRTGKALVSNPESTGKNLKLKNGEPDKTRCTGCKKOKVTOAMH
RKIKDCDEQDELETKDEDSILKLGESADPDRMEDHQELPENINVTETDGEQEA
AKPEPTSKTNGEENPNEPDTGDEKHSKLEPNAPKPSDGEQEAEPNAPKATDGE
NOEAKELTAERKTDDEEHKVADEVEQSKETNVEPEAGEEQSKSEPNAPKTKY
BEKSAKEQTAKLIERKEDMSKTGEEIDKRETSYSSIPETCKVGNAEEDDQVRIKEL
EESDKAEVSTTVLEVDP"

```


2194ACTGTGGAGTCTCTAAAGGCACAGCTTT 2221
 360 isAspProGluGluAlaIleArg.....Hisasp 369
 2222 CAGACGAGGAGTCAACTAAATATATATCGGAAGAGGTCTCCTCCAT 2271
 370 ValIleValSerIleValThraAlaLysLysAspIleLeuValas 386
 2272 GAGTTGCTATCTGAAGAAATTCAAAGCTAAAGATCAACTTCTGTCTGT 2321
 386 nAspHisLeuLeuAsnPhelValArgGluArgThrLeuAspLysArgTrpA 403
 2322 GGAAGAAAGTCTGAGGCATCTAGGGAACAC..... 2352
 403 rgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr 419
 2353CTGGAGAAAGAA 2364
 420 AlaLeuGlnSerAlaAlaGlyLysAspAla..... 429
 2365 AAAACTAAGTTGCGCTCTCTGTGCTGACCTTAAAGAAATTTCCCA 2414
 430AlaLysGlnIleAlaTrpIleLysAspLysLeuLeuH 442
 2415 CCTTGAAGAGGAGATGAAGAAGTTAAGTAAAGTAGAGATGAGCGCTGC 2464
 442 isIleTyrTyrGlnAsnSerIleAspArgLeuValGluArgIle 458
 2465 AT.....AACCTTGATGAAGAAAGGCGCTGGAAGAGAAATT 2502
 459 PheAlaGlnTyrMetValProHisAsnLeuLeuGluThrThrGluArgMetLy 475
 2503 GAATCTCAGCTC.....AAGCATTTAGAG...GAGGATACCAAA 2540
 475 sCysLeuTyrTyrLeuTyrIleThrLeu.....AspLeuAsnAlaValL 490
 2541 AGCAATGAAGCTGCAGGCAAAATTTGGCAGGACTTGTGTGCCATTA 2590
 490 ysAlaLeuAsnGlu.....MetTrp 496
 2591 AACAGCAGAGAAGAGAGGAGTGTCTAAGGTAGTTGATATCTGG 2640
 497 LysCys.....GlnAsnLeuLeuArgHis.....G1 505
 2641 AAGSCCAATATGAGAGTCTCAGCAAAAGATCGCTCAGATTTCTGCCA 2690
 505 nValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla... 520
 2691 CATGCAGGAACAACTGAAGAGCTGAAAAAGACTCATTTCTGATTTATC 2740
 521 ..SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
 2741 AGCAACTGGAGGAGAGGAGCAAACTCTGTATGATAGAGCCAGGCA 2790
 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValle 553
 2791 AGTCAACCAAGTCTTCTCAA.....CTGGAGAAGATAAATCAGTTAGA 2834
 553 uGluAsp..... 555
 2835 GGGTGAATCTGTCTGCACCAACGCCCTGTATCAAAAGAGAGAGACTGAAG 2884
 556AspGluLysIleArg... 560
 2885 AAAAGAAGCTTGTCTGTCTCCTCCTCTGCAGAGAAGAAACTCAAGATT 2934
 561LysGlnLeuGluValLeuValSerProThrCy 571
 2935 GCTTATCAAGGGGAAAGTACGCTCTTCCACCTCGAAGACGCTTAAG 2984
 571 sSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuG 588
 2985 CAATGCAAGCAAGACACCTGGATTGT.....CTAGCCAAAGAACTGT 3025

588 lyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeu 604
 3026 CTGATGAGAAGTATAAGAAAGCAGAAATTTGAAGCCATGGTAAAGTGTTA 3075
 605LeuGluArgIleAlaProValHis..... 612
 3076 AAAGAGCAAAATTCGAAAGAATAGCATCCTCGAATCTGAATGAAAAA 3125
 613IleAspThrGluSerIleSerA 620
 3126 CAGTTTAGCTGTGGTCAAGAGCGAAAAATGTGAATCTGAAAGCTTCTG 3175
 620 laLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspGlu 636
 3176 GAGAGGTGGAGCATCTGAAAGGCAACTAGATGATCCAGTCAGACGAT 3225
 637 AspGluGlyVal..... 640
 3226 AAGGAGCACTAGCTCAGAAAAATATTGAAATCAACACAGCTGATTAATGC 3275
 641ProThrAspGlnAlaIleArgAla.....GlyLeuG 651
 3276 AAAAGAAAAAGCCACATCGGACTTGGCTATTAAATCTGAGATGGCGCTC 3325
 651 luLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAla 667
 3326 AGCTACAGAAAGCAGTGGAT...ACACAAAAAGCGAATTTCAGTGCCTTG 3372
 668 GluThrPheGluSer.....LeuLeuAla..... 675
 3373 CAGAAATCAGCTGCTCGTCTCTCTGTGATCTCTCGCTCTAAAGGAAGGGA 3422
 675 675
 3423 AGTAGAAGACTTAAATAAGGAGGTGCCCTAAGACAAGAGAGAAATACAGC 3472
 676CysLeuLysMetAspAspGluLysValAlaGluAla 687
 3473 AGCAGCAACAAACCATAACTAAGCTCAGGAAGAAGAACCGCATTTGCT 3522
 688 AlaLeu.....GlnIle 691
 3523 GCACCTTAAGATAAAGTCGCCCTGCAGGAAAAAGAGATCAAGCAGCAAGT 3572
 691 ePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgS 708
 3573 CCAGGCCACTAAAGGTGCT.....GAAAGAGAGATGGCTAAATTTGAAT 3616
 708 erAlaLeuLeuProValLeuHisLysSerLysLys..... 720
 3617 CT.....GTAAATTTCTGAAAAATCAAAAAGAAATTTGAATGTTG 3654
 721 ...GlyProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaI 736
 3655 GAACAGACATCCAGACCAAGAGAGAGACCTTTCTTGCATTCAGAGCA 3704
 736 ePheSerSerLys.....GluThrGlnPheAlaGln..... 746
 3705 GCATCAAGCAAACTTGGTGAGTCCCAAGGCTTCAAGCTCTTATTGCAG 3754
 747IlePheGluProLeu 751
 3755 ATCTGAAAAAAGTGCAAAGCAGAGAACTGATTTGTGAGGCCCAA 3804
 752 HisLys.....SerLeuAs 756
 3805 AATAAGCTGCTGAGGCAAAACCGTTGGCCTCAGAAAAAGGCTTCAGTGT 3854
 756 pProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisI 773
 3855 AGAGAGCAGTTAGAGGAATCTCAAGCTTTGAAATTTGAATTCGTAAG 3904


```
1084 rLeuGluSerProLysAspProValLeuPro.....AlaArgPhePheT 1099
      ::::::::::: :::: ||||| :::::::::::
7844 TATAAACACAAATTTCTAGTATTTGCCACTAAAGAAATATTATTC 7893
      :::: ::::
1099 hrGlnProAsp.....Lys 1103
      ::::
7894 AAATGAAGAGAGATGTTACTTAATAATAATAATATTATGAATGAACGAA 7943
      |||
1104 AsnPheSerAsnThrLysAsnTyLeuProProGluMetLysSerPhePh 1120
      ::::::::::||| ||| :::::::::::
7944 AGAATATCAACACGGTGTCTACTACTAATAATACTCTCCAGGATATGA 7993
      |||
1120 eThrProGlyLysProLysThrThrAsnVal...LeuGlyAlaValAsnL 1136
      :::: ::::::::::|||::: ::::
7994 AAATAAATAAATAAAGAAATATATATGAATATTGAACATTTGATA 8043
      |||
1136 ysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerArgMet 1152
      || :::::: :::::::::::
8044 AATTAATAGATCATATAAATAACATATGAATAAATAACAGCAGAAATA 8093
      |||
1153 GluThrValSerAsnAlaSerSerSerSerAsnProSerProGlyAr 1169
      ::::::::::|||::: ::::
8094 TTAATAATTGATGATGCCAAAGAAAAAGTAAAGGAAATAACAGATAATAT 8143
      |||
1169 gIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluA 1186
      |||::: ::::
8144 TAACAAGCTTTTAATCAATACAGAAATTTATAATAAGAAATAATG 8193
      |||
1186 spTyThrMetSerSerProLeuProGlyLysLysSerAspLysArg... 1201
      ||||| |||
8194 GGGTAATT.....AAATCTGCAAAAAATATT 8219
      |||
1202 ...AspAspSerAspLeuValArgSerGluLeuGlu..... 1212
      |||
8220 GTCGATGAAGCTACTTATTAAATGAATTAATAGATAATTTTATTGAA 8269
      |||
1213 .....LysProArgGlyA 1217
      |||
8270 ATTGAATCAATTATTAACTCATATAATAATGATATAAGGATCTTGGTG 8319
      |||
1217 rgLysLysThrProValThrGluGlnGluLysLeuGlyMetAsp 1233
      :::: |||
8320 ATGAAAAATTAATATTAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 8369
      |||
1234 LeuThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerAr 1250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8370 TTGAAAAAGCGCAACAGAAAGAAAGAAAGAAAGAGAGAGAAAGATA 8419
      |||
1250 gLys.....ArgGlyHisThrAlaSerGluSerAspGluGln. 1262
      ||| :::: |||
8420 AAAAGAAAAACAAGAAAGAAAGAAAGCTGGAAGAGAGAAACAAACAC 8469
      |||
1263 ..GlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGlu 1278
      :::: ||||| ||| ||||| ::::
8470 TAAAAAAGAAAGAAAGAAATTAAGAAAAAAGAGCAGGAAAGAAAGAA 8519
      |||
1279 AspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLy 1295
      ::::::::::: ||| :::: ||| |||
8520 CAACAAAAAGAAAGCAAGCTTTAAAGCAACAAAGAAAGCAAGCTACAAA 8569
      |||
1295 sProLeuGlyGlyThrProLysGluGluProThrMetLysThrSerL 1312
      | :::::::::::
8570 A.....GAAGAGAATTAAGAAAGCAAGCAAG 8598
      |||
1312 yLysGlySerLysLysSerGlyProProAlaProGluGluGlu... 1327
      :::: |||
8599 AAAGCTGGAAGAGAGAAACAAAGAAACAACTACAAAAAGAAAGAAATTA 8648
      |||
1328 .....GluGluGluGluArgGlnSer.....GlyAsnTh 1337
      |||
8649 AAAAGCAAGCAAGCAAGCAAGCTACAAAAAGAAAGCAATTAAGAAACA 8698
      |||
1337 rGluGlnLysSerLysGlnHisArgValSerArgAlaGlnG 1354
```

```
|||||::: ::|||:::|||||:::|||||::: |||
8699 AGAACAAAGACGACTACAAAAAGAAAGAAATTAAAAAGACAAGCAAG 8748
      |||
1354 lnArgAlaGluSerProGluSerSerAlaIleGluSerThrGlnSerThr 1370
      ::||| ||| ::::::::::: |||
8749 AAAGGCTGGAAGAGAGAAAAACAACAACACTACAAAAAGAAAGAAATTA 8798
      |||
1371 ProGlnLysGlyArgGlyArgProSerLys 1380
      ::::::::::: ||| ::|||
8799 AAAAGCAAGAAACAAGACGACTACAAAA 8828
      |||
seq_name: gb_ov.XLNUMAPRO

seq_documentation_block:
LOCUS XLNUMAPRO 7321 bp mRNA VRT 20-NOV-1996
DEFINITION X.laevis mRNA for NuMA protein.
ACCESSION Y07624
VERSION Y07624.1 GI:1514670
NUMA gene; NuMA protein.
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 7321)
AUTHORS Merdes A., Ramyar, K., Vechio, J. D. and Cleveland, D. W.
TITLE A complex of NuMA and cytoplasmic dynein is essential for mitotic
spindle assembly
JOURNAL Cell 87 (3), 447-458 (1996)
MEDLINE 97053784
REFERENCE 2 (bases 1 to 7321)
AUTHORS Merdes A.
DIRECT SUBMISSION
SUBMITTED (21-AUG-1996) A. Merdes, Ludwig Institute for Cancer
Research at UCSD, Cell Biology, 9500 Gilman Drive, La Jolla, CA
92093-0660, USA
COMMENT Homologous to human NuMA: Z11583 and Z11584.
FEATURES
Location/Qualifiers
1..7321
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/dev_stage="tadpoles"
196..6957
/gene="NuMA"
196..6957
/gene="NuMA"
/codon_start=1
/product="nuclear/mitotic apparatus protein"
/protein_id="CAA68905.1"
/db_xref="GI:1514671"
/db_xref="SPTREMBL:P70012"
/translation="MTSHSGKMEALLCWNSLKVDEPIERFSQLQDLISLLKVVGLS
GNSETAPVQQDLQRLNCLDSFLQRYCQTGSSAQNVLVQWKILQGENLELVLSKVI
VILFYSTMSSKNPKRSEDIDHKTETELAFILFDNEDALCDLDDKLNFLKRAFP
PSSGNDPSSSDAMTPNVSHRRKSEVFRQLHSVASSVKSLSADAPSPMTEVLHT
POFOMRLKQLSEVREDELELAENRKHLEAKAOISLLOQOENLRILRENOT
ELQEPKLEELREKNESLMRLDTLKOCQDMKADKLLERKNDOLAEENGELSYKVR
DLSNRLAQEALYETTERQELSLSNWQOKNQOSELSCAVGEKYLEEHLNLIQK
TAVEEQAQSRFSEKQGLQETVNLQTSLEITTFQERLNDENARAQHELMCOITT
LKLEISKLSLVHKDEELGTHHKVEERNEKNQLLENFKMLGGNLIGITTOLESKT
KEVDYLRQOQKILCERDSTLNEYCKCKDBNSGVLNKTVKTLQPDHOTLSLSTEK
LASEKEELASKVQDLDAKMLGLIAKCONLDSNDSSQSKSHAATVESLKAQLEQESQ
KIYRKVSNSELYSENSKLQDLISVEESLRLHLEKTKTFAASLDADLKRISH
LEEMKKSSEDEALHNDEERTAGKTESQLHLEETQKNEESQKESQKESQKESQK
KOREERDELSKVVDIWKAKYEQSQKIAQNSCHMOEQTEELKTHSDYVQLEGRS
KVMTEAKASQKESQKESQKESQKESQKESQKESQKESQKESQKESQKESQKESQ
YOGESERLHLETSALNAKQDLCLAKELSDKYLKAEFEAMVKVLEKNSERIASLE
SELKNSLAVVKKERKESQKESQKESQKESQKESQKESQKESQKESQKESQKESQ
ATSDLAIKSEMGAQKQKAVDTHKSEFSAQONELSRSLDLLLALKEGEVERLNEALRQ
EETQQOQTITKLTEETAAALKDKVQKEKEIKQOQVQATGAKEKEMAKLKSIVSEK
SKRIECLQDIONKRLSCIQHQSRLGESQGLQALADLEKKCKEQKELICEAQN
KAAEARTLASEKASVSEROLEGTQALEIETIGKERQACDLQKLELSWAVEBEKTEL
```



```

52 GluLysGluLeuTyrLeuAsnLeuAlaLeuHis.....LeuAlaSerAs 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4734 CAATCAGATTACATTTACGATATGCATCTAGAGTAGTATATATGTTATGA 4783
66 pPhePheLeuLysHis.....ProGlyLysAspVala 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4784 TCTTTTATAAACAATGAAATATAAATCTCTAGCGATGGAAAAAATTTT. 4832
77 rgLeuLeuValAlaCysCysLeuAlaAlaPheArgIle..... 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4833 .....GATATTATAAAGGTTAAAGAATG 4856
91 .....TyrAlaProGln 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4857 ATAAATAAAACCAACAAGTTTCAATGAAGCTATGCAATATGCTAATAA 4906
94 uAlaProTyrThrSerProAspLysLeuLys.....AspIlePheM 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4907 AATGGATGAAAAAATAAGGACATTTATAAANAATAGAAAAATCAACTTTATA 4956
108 etPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGln 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4957 ATTTAATATAATAACATCGTTCATTAAGGGGTAATAATATGAAAAA 5006
125 PheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTy 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5007 GTTAGGAAA.....CAAGCAAG 5023
141 rAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuT 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5024 AAAT.....GCAATTGATGATATAAATAATATACATTTCTAATATT. 5063
158 yrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHis 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5064 ..AAACGATTTAACC.....AAATCTAAAGAACGA 5093
175 MethisMetValAspLeuMetSerSerIleCysGluGlyAspThrVa 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5094 TTAGATGAGATTAAAGAAACCACTTAACATTAAGAGAGAGTGATGTTT 5143
191 lSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisL 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5144 AAATAAT.....GATAAAACCAAAATAGCTTATATACATACAAA 5184
208 yAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu.....LeuLys 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5185 TAAATACGGAAGATAGATCTAATTTTATAATATATATAATATGAAA 5234
224 ArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVa 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5235 CATAAACATAGATACTATCTTGAATAAAGCTATGCGATTATATGAATGATG 5284
240 lLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheA 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5285 ATCAAAATCTGACCAGATGTTATTAATATAGATTCTTGAATATGAACG 5334
257 spLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuSerVal 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5335 AT.....ATATATAAAGATAGAAAGTCTTTTAAATAATATT 5372
274 LeuPro.....GlnLeuGluPheLysLeuLysSerAsnAs 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5373 TTAAGAGAAAAACAGAAATATGGAGGAGCAATATAA...AAAATGAATGA 5419
285 p.....AsnGluGluArgLeuGlnValValLys..... 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5420 AATGATATAATTACGTTAATGAACAGAAAAAGAAATAATAAACAATAAAA 5469
295 .....LeuLeuAlaLysMetPheGly 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5470 AAAATTATGAATAAGAAATTATGGAACATATAAAAAAGAAACAAATGAA 5519
302 AlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrp..... 315

```

```

   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5520 AAAAAAATAAATTTATGGAATCTAATAACAACATCAATACTACTTTAAT 5569
316 .....GlnCysTyrLeuGlyArgPheAsnAspIleHisV 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5570 GGATTCATTCAGATCTATGTTTATATAATGAATATATAAATGATTATAAT 5619
327 alProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsn 343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5620 TAAATGAAATTTTGAA.....AAA 5639
344 HisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerH1 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5640 CATCAAAATATATTGAATGAATATATAATGATTAAATGAA..... 5681
360 sasProGluGluAlaIleArgHisAspValIleValSerIleValThrA 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5682 .....TCATATAATATTATTAAACAAAAATGACTG 5712
377 laAlaLysLysAspIleLeuLeuValAsnAspHisLeu...LeuAsnPhe 392
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5713 AA.....ATTATAAATGATAATTTAGATTATAATGAA 5744
393 ValArgGluArgThrLeuAspLysArgTirPargValArgLysGluAlaMe 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5745 ATAAAGAA.....ATTAAAGAAGTAGCACA 5770
409 tMetGlyLeuAlaGlnIleTyrLysLysTyr.....AlaL 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5771 AACAGAAATGATCAAACTTAATAAAAAAGTTGATGAATTAATAAATATT 5820
421 euGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLys 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5821 TGAATAATTAAGAACAAGAGGACATCGATTAATTGATTATATAAAA 5870
438 AspLysLeuLeuHisIleTyr.....TyrGlnAsnSerI1 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5871 GAAAAAATATTTAACCTTATATAAATGTTTCAGAACACAATAATAAT 5920
449 eAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProH 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5921 AGATGATCTTCTATAATTATATTACAGTTAAAAAACAGTAT..... 5960
466 isAsnLeuGluThrGluArgMetLysCysLeuTyrTyrLeuTyrAla 482
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5961 .....ATTAAACTATTGAGATGCTGAATTTTA..... 5990
483 ThrLeuAsp...LeuAsnAlaVal.....LysAlaLeuAsnG1 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5991 ...TTAGATTCATTGAACACAATAAGAAAAATAAATCAGTAGCAAA 6037
494 uMetTrpLysCys.....GlnAsnLeuLeuArgHisG 505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6038 TCTAGAAATTTGTACTAATAAAGAAGATATAAAAAATTTACTTAAACATG 6087
505 lnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSer 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6088 TTATAAAG..... 6095
522 ValLysAlaIlePheSerLysValMetValIle.....ThrAr 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6096 ...TTGCAAAATTTTTCAGGATTATTGTTAATGCTGATACAAATACGGA 6142
534 gAsnLeuProAsp...ProGlyLysAlaGlnAspPheMetLys..... 547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6143 AATAACCCAGAAAATCCTTTAGAGGATATAATGATTATTATTAATTTACAAT 6192
548 .....LysPheThrGlnValLeuGluAspAsp... 556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6193 TATATTTTGAAGAAACATGAATACATCAACATTGGAAAAATGATTCT 6242
557 .....GluLys 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


14043 CTTCAAGAAATTAAGACTATTCAACTTCGGATGGAAGTTCTCAAGAGT 14092
 722 oProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerS 739
 14093 TTTCAATGAATCAGTCACTGCTGAAGAAATTTACTTCAAGAGCGGTTCAAA 14142
 739 erLysGluThrGln.....PheAlaGlnIlePheGluProLeuHis 752
 14143 ATAGTGAACAACAATGGAACATTTCAAGAAATTTGCAAAATGTTCAAT 14192
 753 LysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeu..... 767
 14193 TGGGAA...GAACAAGCTTTACAGAGAAATCATCAAAATTTGAGAAT 14239
 768ValThrIleGlyHisIleAlaL 775
 14240 TATATGACTGCGATTCACAAGATTCGAAGACACAGTAGGA.....AGCT 14283
 775 euLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThr 791
 14284 TGAAGGCTCCGATGATCAACAGCTCTCCAGATTCATACATTTTCATCA 14333
 792 PheIleValLysAspLeuMetAsn.....AspAr 802
 14334 ACAAGACGACGAGCAAGCTGTATAAAATTTGATGCGACGAGCTGATACGC 14383
 802 gLeuProGlyLysThrThrLysLeuTrpValProAspGluGluValS 819
 14384 TATTTCACATCATCCACCTTCAATGATGGAATCGGATGGAACGCTTTC 14433
 819 erProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
 14434 TTCATAAACTATTCTGAACACGATGCTCTT 14465
 836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuAr 852
 14466 ...TGTAATCTGACAACTCTCGTTTCAAGATCTGAACACAAACAAATGAC 14512
 852 g.....LeuLeuThrThrIleLeuHisSerAspGlyAspLeuT 865
 14513 TCAAGGATTTATCAGAAAGTGGAGAGTGTCAACACAGAACAGAAATTAG 14562
 865 hrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 881
 14563 CTGACGACCTTAATAATAAT...TGTAATGAAGTACTTGAATTTCT 14609
 882 AlaGlySerAla.....IleValLysLeuAlaGlnGluProCyst 895
 14610 GATGACGAAGTTCCATGGTATCTGTTCAAGTCGTAAAGACGACGAATCAT 14659
 895 yrHisGluIleIle.....ThrLeuGluGlnTyrGlnLeuCysAla 908
 14660 CACAAAAGATTTCTTTGGCAAGTGAACCTTGAAAAGCAACACTTCAACA 14709
 909 LeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLy 925
 14710 TTGGCAAGTGGCGATGAA.....ACGAATTC 14735
 925 sLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaI 942
 14736 CATTTATACGGAATTCATCATTCGTTACCGTCCGAGAGAAATTTGAGAAT 14785
 942 leCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAla 958
 14786 TGGTTTCTATCGAGCTTCGGTATGATTGTAGAAAAAGTTTGGAGTATCA 14835
 959 ArgGlnCysLeuValLysAsnIleAsnVal.....ArgArgGluTyr.. 972
 14836 GAGGAAGTTTAAACAACACTCTTAATTTGATTCCTCAAAATGATTTGGAG 14885
 973LeuLysGlnHisAlaAlaValSerGluLysLeuL 984
 14886 CAGTTTGACGATGCCAATTTGCCAAAAAGAAATTTGGTTATCAAAATCTGG 14935

984 euSerLeuLeuPro...GluTyrValValProTyrThrIleHisLeuLeu 999
 14936 CATTTGCTTTTACCTGCGGAGCTCTAATCATGAATCGGTGGCACGATT 14985
 1000 AlaHisAspPro.....AspTyrValLysVa 1008
 14986 CAGGCACCAACCAAGAGGAGCATTGCGAGAGTTGGATTGAAACAAGC 15035
 1008 gInAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValL 1025
 15036 AAGAGAAGCAAAAGTTATGATGGATATTTCAGAAATGC.....GTAT 15076
 1025 euGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg 1041
 15077 TCACACATCTACAGGATACACTGAAATGAATAA..... 15111
 1042 LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAs 1058
 15112CCAGATGA 15119
 1058 pAlaLysMetAsnGlu.....LysLeuTyrThrValCysA 1070
 15120 TGCTGAAATTTACCACATCTCTGCTCTTCTTGTGTACTCTCTCAATCTGTG 15169
 1070 spValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGlu 1086
 15170 ATCTC.....GTCAATGGCTCTTCAAAAT..... 15195
 1087 SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLy 1103
 15196ATTCAAGTTGACAG 15209
 1103 sAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPheP 1120
 15210 TAAATATGACTACTATAGAGCGCAGCTCCAAAAAATGCAGAGACTACAT 15259
 1120 heThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
 15260 TCACAGAAATCCAAT.....GCAGATGCATTTTTCATTG 15291
 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetG1 1153
 15292 GCTCTTCAAGAGCTGGTGAG.....GTTACATCTTCTGTATTTGGAG 15335
 1153 uThrValSerAsnAlaSerSerSerSerSerSerProGlyArgI 1170
 15336 TAGCGTTTCAAACTCAGATGCTCAAAACTACAGTTTCTCACAAGTTGA 15385
 1170 leLysGlyArgLeuAspSerSerGluMet.....Asp 1180
 15386 TTCT.....ATTCTAAACCGAAATGCTATGAAGCGCAGTCAGTGAA 15429
 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLy 1197
 15430 AAGCTGTAATCAGGATCACTCTTG..... 15456
 1197 sSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlyLysP 1214
 15457AAGAAAGATTCATCGAAGCAGTA.....GAGATGAACATTC 15493
 1214 roArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGly 1230
 15494 CGATTTCGCGAAAGAAAT.....CTTCAGAAAACTATTCA 15531
 1231 MetAspAspLeuThrLysLeuValGlnGluLysProLysGlySerG1 1247
 15532 ATTTGATCGATCAAACTCTACTGTTGAATGCAA..... 15564
 1247 nArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnT 1264
 15565GGACAAATTCATCGGAAGACATTTGCAATTCAT 15598


```

215 yAspLeuAlaLysAlaLeu.....Leu 222
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12376 TAAATCTCAGAGAACTTTAGATCAAAACGGTCAAAATGACGTCTCGCTT 12425
223 LysArgThrAlaGlnAlaIleGluProTyrIleThrThr..... 235
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12426 ACCGATCAGATGAGAGATCGAAGTTAAAGTTTCCAAATATATATCTCTTT 12475
236 .....PhePheAsnGlnValLeuMetLeuG 244
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12476 GATTTCTCCGAAATCTTGGAGATCTGATATCACAAGCGTTCAACTTG 12525
244 LylsThrSerIleSerAspLeuSerGluHisValPheAspIleLeu 260
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12526 CTCACCAATATGCTGTATGAGACAGAAAAATCATTTGAA..... 12566
261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnL 277
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12567 .....ATTCCAAGAACAATCTCTTAAACAGAGTTGTGCCAGA 12604
277 euGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVal 293
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12605 AACATCGGAAGAATCATTAATCATCTCTGGAAGACATCTCAACTTGCTG 12654
294 LysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerG 310
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12655 AAGAAGCTCGTCAACTATTT...CTGAAAGCTATCAATGTTGCAATCA 12701
310 nAsnLys...ProLeuTyrGlnCysTyrLeuGlyArgPheAsnAspIleH 326
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12702 GAATTTGTAGTCTCTGCTGCAAAACAAGTTA.....GTAC 12736
326 isValPro..IleArgLeuGluCysValLysPheAlaSer..HisCysLeuM 342
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12737 ATCTCTCACATTTGGATTACCGCGGAAGATTTTCAATCAAAATTCGAAA 12786
342 etAsnHisProAspLeuAlaLysAspLeu..... 351
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12787 TTATATTCGGTATCTGACGAAGAGATGTCGTCGAAGCGGCTTTCTCAGTA 12836
352 .....ThrGluTyrLeuLysValArgSerHisAspPr 362
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12837 TCAGAGAAACATTAATACATTAATTTCTGCAATTCGGAATCGGATG 12886
362 oGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 379
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12887 GAGTCAGATCAAGTTATCTTCCAAGCAAAATCTATATATCTGCTAACA 12936
379 yLysAspIleLeuLeuValAsn...AspHisLeuLeuAsnPheVal... 393
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12937 TAAATCTCTCGCAACATCTAATCTAGACACCCCTCATTTGGACAACGTATG 12986
394 .....ArgGluArgThrLeuAspLysArgTr 402
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12987 CCAAGAGAACGATTCGGAGAGATGAGCTTCATTCTCTCAAGCAAG 13036
402 pArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle..... 415
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13037 CAGTGTGAGAAATGGAGAACACTTCAAAATTCAAACAATTTATTTTGATA 13086
416 .....TyrLysLysTyr..... 419
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13087 CAGTTGTAGAAAGTTTGGATCCGATCGAAGAAAGGCATCAATAATAAT 13136
420 .....AlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAl 434
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13137 CTCATTTGCTCTGCTGACTCTGTATTTGAAGCGACGCTATCCAAGTTGC 13186
434 a..TrpIleLysAspLysLeuLeuHisIleTyr..... 444
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13187 TTTCTTTGAAGAACTACCTGATTCACCTCGATTCACAGCTTCAATCACTC 13236
445 .....TyrGlnAsnSerIleAspAspArgLeuLeuValGluArg 457

```

```

13237 CAGTCTTCAAGATTCAAGAACATGTCCTCGACCGACTG.....AACAG 13280
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
457 gIlePheAlaGln.....TyrMetValProHisA 467
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13281 GTTGTCAAGGATTCGGTCAAGGAAGAACATCCAAATGAACCGTGA 13330
467 snLeuGluThrThrGluArgMetLysCys...LeuTyrTyrLeuTyrAla 482
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13331 AGTTG...TGAATAGAAAAATTGAGGTGCTGAAGAGTATTTTGAATTGTT 13377
483 ThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys.. 498
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13378 ATGCTGTGCTGAATGCCATAAGAGTTTGGAGAGAAGCCTGGAGAAGCCT 13427
499 .....GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspL 512
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13428 TCTCAGCAGATGAACACTTGTGCGACCCGAG.....GAGTTACACAAGT 13471
512 euIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13472 TGTTCAGAATGCGTTT.....GCTGTAACGGAACAATCGTATTTCCAAG 13514
529 ValMetValIle.....ThrArgAsnLeuProAspProGlyLysAl 542
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13515 ATTCGGAAGGTTTAGGATCAATGCAAAACTTTGAAATACCAAGAAGCAGT 13564
542 aGlnAspPheMetLysLysPheThrGlnValLeuGlu.....A 555
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13565 AAAAGAAACCATTTCAAAAAATTTGAAATTTGGAATTTGCCACCAGTCG 13614
555 spAspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCys 571
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13615 ACACATTATTGTTATCAAAATATAAAGAAATCACATGTCAGCAAGACAGTC 13664
572 SerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuG 588
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13665 GAA.....GAACAAGAATTTGGTTCAGTTCAATGTT 13693
588 yAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuL 605
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13694 TAATTTCAATCGGCACACTCAATTCGAAATTTGAAATGTCACAAAATGTTT 13743
605 euGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeu 621
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13744 CAAAATGCAATTTGAAGAGGAACAGTCGTTCTTCCCAACCTTGTAGTATT 13793
622 IleLysGlnValAsnLysSerIleAspGlyThrAlaAspGluAspG 638
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13794 TTAGCTTCTCGCAATCTTGTACAATT..... 13820
638 uGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysv 655
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13821 ....GTTCTACTGAGAGCAGACAACCTTTCCGATTTCAGAAAATATTCTCA 13866
655 aLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGlu 671
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13867 ATTTCTCGATACAACAGCAAGAAATCCGAGCATCAGTCCACAGCCGT 13916
672 SerLeuLeuAlaCysLeuLysMetAspGluLysValAlaAlaAla 688
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13917 .....GATGAAATAAACATTTCTAAGGCATTA 13942
688 aLeuGlnIlePhe.....LysAsnThrGlySer..... 697
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13943 TCTGCAAAACATTTGCCATTCAAAAGGCATCTTCAGAAAGTGACAGAGTAA 13992
698 .....LysIleGluGluAspPheProHisIleArgSerAlaLeu 710
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13993 TAGAAATCATTAATAAAGTTCTCAAGGTTTCCATTTGGTTGCACAGCAATG 14042
711 LeuPro.....ValLeuHisHisLysSerLysLysGlyPr 722
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```

795 ysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu 811
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12323 GTGATCACTGGAGAGCTCTG.....AAGAAACGAAGAAGT 12363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
812 TrpValProAspGluGluValSerProGlu..... 821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12364 TGGTCCCTCCAGAGGAATCTTTATTAGAACAGCTAGATTATCGAATTC 12413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
822 .....ThrMetValLysIleGlnAlaIleLysMetM 832
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12414 TCGATTGAGGATTATCTCCCAAAATAAATCTATTGATGATCAAAATAC 12463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 etValArgTrpLeuLeuGlyMetLysAsnHisSerLysSerGlyThr 848
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12464 AAATCTACAGCTGCGGCAAGAGTCAATAATTCGACAAACGGACCT 12513
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 SerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyLeuLys 865
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12514 GGTGTGAATAATATTTAATTACACTACGTGCGGAAAGGGATATCTTGA 12563
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
865 rGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12564 TACAAGAGTGCAGTGGCTGAAAGAGATGCAAAATCTTGAGACAAAAA 12613
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
882 laGlySerAlaIleValLysLeu.....AlaGlnGluProCysTyrHis 896
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12614 TTTCTTTGATGATGTTGCAATTACAAAGTCTCGTACTAGCTAGATAAT 12663
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
897 GluIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAs 913
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12664 TCAAGAGTTGAAAGGAAAAATCATCTCTCCATTTTCAACAGCATGACGA 12713
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
913 pGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyL 930
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12714 C.....ATTATGAGAAAT 12727
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
930 euSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12728 TAATCAATTAATCTA..... 12744
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
947 AlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuVal 963
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12745 .....TTAAGAGAAAGTAACATACATTGCGGAATGAGCTGGA 12782
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
963 lLysAsnIleAsnValArgArgGluTyr.....LeuL 974
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12783 AAACAACAATAAAGAAGAAAGGAACTGCAATCTGAATTAGATAAATGA 12832
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
974 ysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyr 990
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12833 AGCAAAATGTTGCGCTATCGAGTCCGAATTCAGACGCTTG..... 12873
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
991 ValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLy 1007
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12874 .....AAATATTCTAT 12884
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1007 svalGlnAspIleGlnLeuLysAspValLysGluCysLeuTrpPheV 1024
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12885 GCAAGAAAAGAGCAAGAGCTCAAAATTAGCTAAAGAA..... 12921
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 alLeuGluIle.....LeuMetAlaLys 1031
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12922 .....GAGTTCATCTGTTGGAAAAGCGCTCACAGACATATTGGAGAA 12966
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMet.....ValG 1045
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12967 CATGAACAATTGAGCTCAAGCGATTATGAGAAGCTAGAAACGGAGATAGA 13016
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1045 uasnIleLys.....GlnThrLysAspAlaGlnGlyProAsp...A 1058
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13017 AAATTTGAAGGAGGAACCTAGAAAAATAAGGCGTCAAGGCGGAAGCCG 13066
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 spAlaLysMetAsn.....GluLysLeuTyrThr 1067
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13067 AGGAAAAATTTAACAGCGCTGAGAGACAACGCGCAAGAGATTAAAAACA 13116
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1068 ValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSe 1084
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13117 .....TCAAAACTCTCACAGGACTCATTTGACTGAACA 13148
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1084 rLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnP 1101
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13149 AGTAAATAGTCTAAGGATGCA..... 13170
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1101 roAspLysAsn...PheSerAsnThrLysAsnTyrLeuProProGluMet 1116
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13171 .....AGAACGCTGTTGAAAAATTCCTGAGTGAGCAAAACGCCGAGATC 13215
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAl 1133
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13216 GAAGAGTTACAAATGCAAAAGTAGCACAAAGTAACACCAAGTTAGAGC 13265
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1133 aValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13266 AATAAGAAAAATTACAGAAGACGCAAAAGGTTCAAGAGAGCTTCAAG 13315
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1150 erArgMetGluThrValSerAsnAlaSerSerSerSerSerSerSer 1166
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13316 CCAAGTTAGAA.....GAAAGTACGACTTCTTAC 13344
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1167 ProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerG 1183
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13345 GAATCTACGATAAAGCGCTTA..... 13365
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1183 uAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspL 1200
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13366 .AATGAAGAGATTACA..... 13380
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1200 ysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGly 1216
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13381 .....ACATTAAGAAGAAATAATGAACAAAGGCAA 13413
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1217 ...ArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAs 1232
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13414 ATCCAGCAACAGATTACAGGTACATCTGCAAAATGAACAA.....AA 13454
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1232 pAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnArgS 1249
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13455 TGACTTATCAACATAGTTGAGTCTATGAAA..... 13485
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 erArgLysArgGlyHisThrAlaSerGluSerAspGluGlnTrpPro 1265
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13486 .....AAGTCTTTTGAAGAAGATAAAATCAAAATTCATC 13518
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1266 GluGluLys.....ArgLeuLysGluAspIleLeuGluAsnGluAspG 1280
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13519 AAAGAAAAAACCCCAAGAAGTTAATGAAAAATACTCGAGGCCCAAGAAAG 13568
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1280 uGlnAsnSerProPro.....LysLysGlyLysA 1290
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13569 GCTAAATCAACCTTCCCAATATCAATATGGAGGAGATTAAAAAAAATGGG 13618
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1290 rgGlyArgProProLysProLeuGlyGlyThrProLysGluGluPro 1306
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13619 AATCTGAGCAGCAAGAGAGATATCTCAAAAGATTCGCGAAGCTGAGGAA 13668
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1307 ThrMetLys.....ThrSerLysLysGlySerLys.. 1316
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13669 GCCCTCAAAAAGCGAATACAGATTACCCACTGAGGAGAAAAATTAAGAT 13718
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1317 .....LysLysSerGlyProProAlaProGluGluGluGluG 1330
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13719 AATCGAACGAAGAAGAGGGAANTTGGAAAAAGAGTTTGAAGAAAAGGTTG 13768
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1330 luGluArg.....GlnSerGly..... 1335
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


[illegible]

302 alYsnAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrIL 319
 ::::||||| ::::||||| :||
2785 GAAGAGTCAACAGCTAAATGAGGTAAA..... 2811

319 euGIyArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335
 ::: ::: ::: ::: ::: :||| ::: ::
2812 GAGGCTATATACTTTAAGAGACTGAAAAAATGGAT 2847

336 pheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuTh 352
 :: :|||:: :|||:: :||| :||| :||| :||| :|||
2848 CTAGAATCAAGAAATT.....GAGGAACACTACAGAAAGAGCTTGA 2885

352 rGUlTyrrLysValArgSerHisAspProGluGluAlaIleArgHisA 369
 ||| | |||||:: :||| :||| :||| :||| :||| :||| :|||
2886 AGAA...TTGAANACT...TCTGTTCACCAGGAATGCCGTATACAGCA 2929

369 spValIleValSerIleValThrAlaAlaLysLysAspIle...LeuLeu 384
 :||| :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||
2930 ATGTAACATATAAACACAGTTAACCGAAACTTAGAGAGACCTCGAATCTCAA 2979

385 ValAsnAsp.....HisLeuLeuAsnPheValArgGluArgThrLe 398
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
2980 GTACAAGACTTTGCAAACTCGTATCTCCGAATTACTAGGAGAGTCTACT.. 3027

398 uASpLysArgTrpArgValArgLysLysGluAlaMetMetLysLeuAlaGlnI 415
 ::::~::~ :|||:||||| :|||:||||| :|||:||||| :|||
3028 GAAATAATGTCACCTTTTAAACAAGGAG.....ATACAGGACC 3064

415 leTYr...LysLSyTYrAlaLeuGlnSerAlaLaGLyLysAspAlaAla 430
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3065 TGTATGACAGCAAGACGCACATATCCATTAAGCTTGGAAGAGAAAATCA 3114

431 LysGlnIle...AlaTpIleLysAspLysLeuLeuHisIleTyrrG1 446
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3115 TCGAGAATATTGGCAGGAGACGATTTAAACTACTT.....TC 3152

446 nAsnSerIleAspArgLeuLeuValGluArgIlePheAlaGlnTyrm 463
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3153 GAATACGPTAGAT..... 3165

463 etValProHisAsnLeuGluThrThrGluArgMetLysCysLeutyrrTy 479
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3166 CTAACTTAAGCTGAGAACCAACCATCGCCAAAGGTTTGATTT 3210

480 LeutyrrAlaThrLeu.....AspLeuAsnAlaValLysAlaLeuAs 493
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3211 TTACAGAATACTATTMTTAAACAAGATTTCCAAAACACACAGACACTTAA 3260

493 nGluMetTrpLysCysGlnAsnLeuLeuArgRHISGlnValLysAspLeu 510
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3261 TGAATACGTTTCCCTGTAAATCTAAGTTTAAGCATTTGTGAAACAGAAATTAT 3310

510 euASPLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePhe 526
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3311 TGAACCTG.....AAGAACAACAGAAATTAAGAGTTTCATTTA 3348

527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGl 543
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3349 GAAAAGAACTTTGAAACAAGAACTGAATTAACACTCTCCCTCGAAAG.... 3393

543 nAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleA 560
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3394 .CACAGTTTTCGCATCATGGTAACTCAATTACAACTTTTACAAAGAGACC 3442

560 rgLysGlnLeuGluValSerProThrCysSercysLys..... 574
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3443 GTGAAGATCTATTGGAAGAGACTAGGAAATCATGTCAAAGAGAAATAGAT 3492

575 GlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro... 590
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||
3493 GAAC TTGAAGATGCTCTCAGCAACTTTAAAAGAGGAACACTTCTCAAAGA 3542

591 LysGlnProThrAsnProPheLeuGluMetI 601

4855 ATCCAGCAACAGTTACAAAGCTACATCTGCAAAATGAACAA.....AA 4895
1232 pAspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgS 1249
4896 TGACTTATCTACATAGTTGAGTCTATGAAA..... 4926
1249 erArgLysArgGlyHisThrAlaSerGlySerAspGluGlnGlnTrpPro 1265
4927AAGTCTTTTGAAGAAGATAAAATCAATTCATC 4959
1266 GluGluLys.....ArgLeuLysGluAspIleLeuGluAsnGluaspG1 1280
4960 AAGAAAAACCCCAAGAGTAAATGAAAAAATATCTCGAGGCCCAAGAAA 5009
1280 uGlnAsnSerProPro.....LysLysGlyLysA 1290
5010 GCTAAATCAACCTTCCAAATATCAATATGAGGAGATTAATAAAATATGG 5059
1290 rGlyArgProProLysProLysProLysGlyGlyThrProLysGluGluPro 1306
5060 AATCTGAGCAGCAGACAGAGTATCTCAAAAGATTCGCGAAGCTGAGAA 5109
1307 ThrMetLys.....ThrSerLysLysGlySerLys... 1316
5110 GCCTCAAAAGCAATACAGATACCACTGAGGAGAAAATTAATAGAT 5159
1317LysLysSerGlyProProAlaProGluGluGluGluGluG 1330
5160 AATCGAAGAAAGAGGAGGAAATGGAAAAAGAGTTGAAAGAAAGTTG 5209
1330 luGluArg.....GlnSerGly..... 1335
5210 AGGAGAGATAAATCAATGACATCTCGAGAAATAGACGTGGTCTT 5259
1336AsnThrGluGlnLysSerLysSerLysGlnH1 1346
5260 CGAAAACAGCTAGAACGTAAGTTCAGAGAAACAAAGGAATTTGAAAA 5309
1346 sArgValSerArgArgAlaGlnGlnArgAlaGluSer....ProGluSerS 1362
5310 CGAGTATACAAAATTAATCAAGAAAGTAACAAAGATGTACCACATCAA 5359
1362 erAlaIleGluSerThrGln 1368
5360 GTCATATCTCAGATGAGAA 5379
seq_name: gb_p14:SCYKR095W

seq_documentation_block:
LOCUS SCYKR095W 8050 bp DNA PLN 18-MAR-1998
DEFINITION S.cerevisiae chromosome XI reading frame ORF YKR095W.
ACCESSION Z28320 Y13137
VERSION Z28320.1 GI:486586
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 8050)
AUTHORS Baladron,V., Ballesta,J.P.G., Bou,G., del Rey,F., Esteban,P.F.,
Garcia-Cantalejo,J.M., Garcia-Ramirez,J.J., Gonzalez,A.,
Jimenez,A., Revuelta,J.L. and Santos,M.A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 8050)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1994) Data collected by MIPS on behalf of the
European yeast chromosome XI sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mpmic.mips.biochem.mpg.de
FEATURES
source
1..8050
/organism="Saccharomyces cerevisiae"

gene
exon

CDS

/db_xref="taxon:4932"
/chromosome="XI"
complement(1..8)
/gene="RPL40B"
complement(1..8)
/gene="RPL40B"
/number=1
complement(<1..8)
/gene="RPL40B"
/note="ORF YKR094C"
/codon_start=1
/protein_id="CAA82175.1"
/db_xref="GI:4388588"
/translation="MQ"
1063..6690
/gene="MLP1"
1063..6690
/note="ORF YKR095W"
/codon_start=1
/protein_id="CAA82174.1"
/db_xref="GI:486587"
/translation="MSRDHDTPMESIQNGENSDELRNIAISFFGCSLEQVKSFGDGVVK
HLNDKLOFNEKLSNKLKVTYFDELKASSLKIDGLKTEMENVIRENDKIERLVRV
FVKPESVENEKMLSELEFVKRLDLDTEKKETQSNQORTLKLIDRLRLKTELVRV
ENRNSSECKLRSTIMDETQOQYIINDLNSTELERKTOELTLLQSNLQLEKLL
RSKNEQYLSRQTDKVLIDINELNRDNDFOMERTNNDVLKQKNLSLSQLEKLL
EIKGLSDLSNKEQFSAEMSLKQRLVDLSQNAVEELNSRELNTAKVIADDSK
KQTPENEDLLKELQTLKLAQCEKCLLSITDEADENELSAKSSDSIFLKKQ
LIKERTKHLNQIETFIVELEHKVPINISFKERTDMLNENLNAALLILEHTSEKN
AKVKELNAKNOKLVCEENLOTLTQRDLQCRQIOYLITNSVSDSKGPKRKEIQF
IQNTMOEDDSTITESDSQKVTYRLEVEFNIIQLEKNAELLKVYRNLDKLESEKK
SQLOKIESETVNEAKAIIITLKSEKMDLESRIEELQLEELKTSVPNEASISNV
TIKQLETKRDLSEQVQDLQTRISQITRESTENSLNKEIQDLYDSKSDISIKLGE
KSSRIAEERFKLSNTLIDTKAENDQRLKRFYDQNTILKQDSKTHETLNEVSDKS
KLSIVTELLNLKEQKLVHLEKMLKQELNKLSPKDSLRIMVQLOTLOKREYDCLL
EETRKSCKDIKLEDALSELKQKTHETQKDHHTKOLEEDNNSNIWQNKIEALKKDYE
SVITSVDSKQTDIEKLOYKVKLEKEIEEDKIRLHTYVMDETINDLSIRKLEKSKI
NLTDAYSQIKYKDIYETISQLOQTNKLDSEFPDFTNQIKNLDEKTSLEDKISLL
KEQMFNLNDELQKQWKEKADFKRISILQNNKEVEAVKSEYKSLKIQNDLD
QOTIYANTAQNNEQELKHADVSKTISLRELQHTYKGVKTLNLSRDLQENALKEN
EKSWSQKESLEQLDLSNRIEDLSQNKLLYDQIYITADKENVNNTNGPGLNI
LITLRERDILDTKVTAERDAKMLRQISLMDVELQDARTKLDNSRVEKENHSIIQ
QHDDIMEKLNOLLNRESNITLNRLENNNKKKLOSELDKLKNVAPIESELTKALK
YSQKEQELKLAKEEVHWKRSODILEKHQOLSSDYKLESIELENKELENKER
QGAEEERFNLRQQRKLTSLQSDSLTEQVNSLRDKNVNSLESEARIEEL
QNAKVAOQNNLOELKQEDAEKASRELQAKLEESTTSYESTINGLNEEITTLKEEI
EKQRIQOQLQATSANEQNDLSNIVESMKKSFEDKIKFKEKTEVNEKILEAQERL
NOPSINMEEKKWESEHEQVSOKIREAEALKRIRLPTEEKINKTIERKKBELE
KEFEKVEERIKSMEQSGEIDVVLKOLEAKVOEKQLENEYKKLOPELKDVPHESS
HISDDERKLAIESRLREEFNNELOAIKKKSFDEGQQAQMMKTLERKLAKESQ
LSETKQSAESPSPKSVNNYONPLGLPRKIEENSNSFPNPLLSGKLLKLNKSSGGF
NPFTSPSPKHLQNDNRKRESLANKTPDTPHLEPSFNIPASRGLSSSSTLSSTDNDE
ELTSNNPAQKDSNRNVQSEDETEKKKEGEPVKRGEATEEQTKSNKRPIDEVGLKND
EDDTTENINESKIKTEDEEKEKETKVDNDSI"

BASE COUNT 3194 a 1321 c 1540 g 1995 t
ORIGIN

alignment_scores:
Quality: 235.50 Length: 1540
Ratio: 0.325 Gaps: 72
Percent Similarity: 47.078 Percent Identity: 19.545

alignment_block:
US-09-512-581-2 x SCYKR095W ..
Align seg 1/1 to: SCYKR095W from: 1 to: 8050
19 ValLysGluLysSerAspLysLysSerLysGlu..... 29
1894 ATCAAGGGTCTATCCGACTCCCTAAATTCGAAAGCAAGAAATTTCTGTC 1943


```

701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisIle 717
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3367 GAATCTTCAAGGACTTTACTAACAGATTAAACCACTAACTGATGAAAA 3416
717 sSer..... 718
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3417 AACTAGTTGGAGATAGATTTCCTCTCTAAAGAGCAAAATGTTTAATT 3466
719 .....LysLysGlyProProAlaAla... 726
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3467 TGAATAANTCAGCTAGATTGCGAGAAAAAGGATGCAAAAAAGAAAGCT 3516
727 .....LysTyrAlaIleHisCys.....IleHi 734
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3517 GACTTTAAGAAAAGGATATCAATTTTACAGACAATAATAAAGAAAGTCGA 3566
734 salalilePheSerLysGluThrGlnPheAlaGlnIle..... 747
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3567 AGCTGTTAAGTCGGAATATGAATCGAAGTTATCAAAAATCCAAAAGGACC 3616
748 .....PheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHis 762
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3617 TTGATCAACAACACTATATATGCTAATACTGCGCAAAACAACACTATGACAA 3666
763 LeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAs 779
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3667 GAACACAGAAAACATGCAGATGTTTCTAAGACGATTAGTGAATTAAGAGA 3716
779 pGlnPheAlaAlaProTyrLysSerTrpValAlaThrPhe...IleValL 795
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3717 GCATTTACATACG...TCAAAAGGTCAGTTAAGACCCTGAACTTATCGC 3763
795 ysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu 811
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3764 GCGATCAACTCGAGATGCTCG.....AAGAAAACGAAGAGAT 3804
812 TrpValProAspGluGluValSerProGlu..... 821
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3805 TGGTCTCCAGAGAAATCTTTATAGAACACAGCTAGATTTTATCGAATTC 3854
822 .....ThrMetValLysIleGlnAlaIleLysMetM 832
3855 TCGTATTGAGGATTTATCTCCCAAAATAACTATTGTATGATCAAAATC 3904
832 etValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThr 848
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3905 AAATCTACACAGCTCGGCAAGAAGTCAATAATTCGACAAACGGACCT 3954
849 SerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuTh 865
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3955 GGTTCGAATAATATTATTAATACACTACGTCGCGAAAGGGATATTCTTGA 4004
865 rGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaA 882
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4005 TACAAAAGTCGCGGTGGAAGAGATGCAAAAATGTTGACACAAAAA 4054
882 laGlySerAlaIleValLysLeu.....AlaGlnGluProCysTyrHis 896
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4055 TTTCTTTGATGATGTTGAATTAAGATGCTCGTACTAAGCTAGATAAT 4104
897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAs 913
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4105 TCAAGAGTTGAAAGGAAAATCATTTTCCATTTATTCACACAGCATGACGA 4154
913 pGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyL 930
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4155 C.....ATTATGGAGAAAT 4168
930 euSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4169 TAAATCAATTAATCTA..... 4185
947 AlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuVa 963

```

```

4186 .....TTAAGAGAAAGTAAACATAACATTCCGGAATCAGCTGCA 4223
963 lLysAsnIleAsnValArgArgGluTyr.....LeuL 974
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4224 AAACAACAATAACAAGAGAGAACTGCAATCTGAANTAGATAAATTGA 4273
974 ysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyr 990
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4274 AGCAAAATGTTGCGCTATCGAGTCGGAATTCACAGCCTTG..... 4314
991 ValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVally 1007
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4315 .....AAATATCTTAT 4325
1007 sValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheV 1024
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4326 GCAAGAAAAAGAGCAAGAGCTCAATTTAGCTAAAGAA..... 4362
1024 aLeuGluIle.....LeuMetAlaLys 1031
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4363 .....GAGTTTCATCGTTGAAAAAGCGCTCACAAACATATTGGAGAA 4407
1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMet.....ValG 1045
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4408 CATGAACAATTCAGCTCAAGCGATTATGAGAAGCTAGAAAGCGAGATAGA 4457
1045 uAsnIleLys.....GlnThrLysAspAlaGlnGlyProAsp...A 1058
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4458 AAATTTGAAGGAGGAACTAGAAAAAAGGAGCGTCAAGGCGGAGCGG 4507
1058 spAlaLysMetAsn.....GluLysLeuTyrThr 1067
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4508 AGGAAAAATTTAACAGCGTGAAGACAGCAAGCGCAAGAGAGATTAACAA 4557
1068 ValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSe 1084
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4558 .....TCAAAACTCTCACAGGACTCATTCAGCTGAACA 4589
1084 rLeuGluSerProLysAspProValLeuProAlaArgPheThrGlnP 1101
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4590 AGTAATAGCTAAGGGATGCA..... 4611
1101 roAspLysAsn...PheSerAsnThrLysAsnTyrLeuProGluMet 1116
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4612 .....AAGAACGTGTTGAAAAATTCCTTGAGTGAGGCAAAACGCAATC 4656
1117 LysSerPheThrProGlyLysProLysThrThrAsnValLeuGlyAl 1133
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4657 GAAGAGTTACAAAATGCAAAAGTAGACAAAGTAAACACAGTTAGAAGC 4706
1133 aValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4707 ATATAGAAAATTCAGAAGAGCGCAGAAAAGGCTTCAAGAGAGCTTCAAG 4756
1150 exArgMetGluThrValSerAsnAlaSerSerSerSerSerSerSer 1166
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4757 CCAAGTTAGAA.....GAAAGTACGACTTCTTAC 4785
1167 ProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGl 1183
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4786 GAATCTACGATAACGGCTTA..... 4806
1183 uAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspL 1200
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4807 .AATCAAGAGATTACA..... 4821
1200 ysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGly 1216
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4822 .....ACATTTAAAGAGAAATTCAAAAACAAAGCAA 4854
1217 ...ArgLysLysThrProValThrGluGlnGluLysLeuGlyMetAs 1232
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```


1763 TCGTAGAGTTGGACATAAAGTCCCATTTATAAAGTCTTTCAAA..... 1806
 145 PheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLe 161
 1807 ...GAAAGAACTGCATGATGTTGGAAACGAATTAATAACGCTCATGTT 1853
 161 uPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetV 178
 1854 ACTAGACATACATCGAACGAGAAGATGCAAGGTTAAGGAATTAATG 1903
 178 alAspLeuMetSerSerIleIleCysGluGlyAsp.....ThrValSer 192
 1904 CCAAAACCAAAAGCTAGTGGAATGTGAAATGATCTTCAAACTTTAACT 1953
 193 GlnGluLeuLeuAspThrValLeuValAsnLeuValProIleHisLysAs 209
 1954 AAACAGCTCTCGAT.....CTATGCCGTCAAAATACAATACCTTTTAAT 1997
 209 nLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrA 226
 1998 TACCAATTCGTGTTCTTAATGACTCGAAGGACCCCTTACGTAAGAGAAA 2047
 226 laglnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMet 242
 2048 TTCAA.....TTTATTCAAAACATTATGCAG 2073
 243 LeuGlyLysThrSerIle.....SerAspLeuSerGluHisValPheAs 257
 2074 GAAGACGATAGTACTATPCAGACAGATGACTCTCAAAAAGTCGTAACGTA 2123
 257 pLeuIleLeuGluLeuTyrAsnIle.....AspSerHisL 269
 2124 AGACTAGTTGAATTCAAAACATATTCAATTACAGAAAAAATGCAG 2173
 269 euLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAsp 285
 2174 AACTTTTGAAGTAGTAAGAACTTACGCCGATAAGTTGGAATCGAAAGAA 2223
 286 AsnGluCluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAl 302
 2224 AGAAATCTAAACAAAGTCTTCAGAAATC..... 2253
 302 alYsAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrL 319
 2254 ...GAAAGTGAACACAGTAAATGAGGCTAAA..... 2280
 319 euGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335
 2281GAGGCTATAATAACTTTAAAGAGTGAAAAAATGGAT 2316
 336 PheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuTh 352
 2317 CTAGAATCAAGAATT.....GAGGAACACTACAGAAAGAGCTTGA 2354
 352 rGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369
 2355 AGAA...TTGAAAACT...TCTGTGCCACAGAGATGCGCTACACGCA 2398
 369 spValIleValSerIleValThrAlaAlaLysLysAspIle...LeuLeu 384
 2399 ATGTAACATATAAACAGTTAACCGGAACTTAAGAGACCTCGAATCTCAA 2448
 385 ValAsnAsp.....HisLeuLeuAsnPheValArgGluArgThrLe 398
 2449 GTACAGAGCTTGCAAACTCGTATCTCGCAAAATTACTAGGGAGTCTACT.. 2496
 398 uAspLysArgTyrPArgValArgLysGluAlaMetMetGlyLeuAlaGlnI 415
 2497 .GAAATATGTCACCTTTTAACAGAGGAG.....ATACAGGACC 2533
 415 leTyr...LysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAla 430
 2534 TGTATGACAGCAAGAGGACATATCCATTAAAGCTTGGAAAGGAAAAATCA 2583

431 LysGlnIle...AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGl 446
 2584 TCAGAAATATTGCAGAGCAAGCACTTTAAACTACTT.....TC 2621
 446 nAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrM 463
 2622 GAATACGTTAGAT..... 2634
 463 etValProHisAsnLeuGluThrThrGluArgMetLysCysLysLeuTyrTyr 479
 2635CTAACTAAGCTGAGACGACCACTGCGCAAAAGGTTTGATTAT 2679
 480 LeuTyrAlaThrLeu.....AspLeuAsnAlaValLysAlaLeuAs 493
 2680 TTACAGAACTACTATTTAAACCAAGATTCCAAAACACACAGACACTTAA 2729
 493 nGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuL 510
 2730 TGAATACGTTTCTCTGTAATCTAAGTTAAGCAATTGTTGAAACAGAAATTAT 2779
 510 euAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePhe 526
 2780 TGAACCTG.....AAAGAACAGACAGAAATTAAGAGTTCAATTA 2817
 527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGl 543
 2818 GAAAGAACTTGAACCAAGAACTGAATAAATCTCCCTCGAAAG..... 2862
 543 nAspPheMetLysLysPheThrGlnValLeuGluAspGluLysIleA 560
 2863 .GACAGTTTACGCATCATGCTAATCAATTACAACTTTACAAAGGAGC 2911
 560 rgLysGlnLeuGluValLeuValSerProThrCysSerCysLys..... 574
 2912 GTGAAGATCTATTGGAAGAGACTAGGAAATCATGTCAAAAGAAATAGAT 2961
 575 GlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro... 590
 2962 GAACTTGAAGATGCTCTCAGCACTTTAAAGAGAACTTCTCAAAAGA 3011
 591LysGlnProThrAsnProPheLeuGluMetI 601
 3012 CCATCATATCAACAGCTGCGAAGAGACACAATTCAAATATAGAAATGGT 3061
 601 leLysPheLeuLeuGluArgIleAlaProValHis..... 612
 3062 ACCAAAATAAAATCGAAGCTTTGAAGAAAGATTATGAATCAGTAATAACT 3111
 613IleAspThrGluSerIleSerAlaLeuIleLy 623
 3112 TCTGTAGATAGTAGCAAACTGACATTGAGAAATTTACAATATAAGTCAA 3161
 623 sGlnValAsnLysSerIleAsp.....GlyThrAlaA 634
 3162 ATCACTAGAAAAGAAATCGAGGAGGACAGATTCGTTTACATACTTATA 3211
 634 spAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
 3212 ATGTTATGATGAAACAATT...AACGATGATTCCCTACGCAAGGAGTTG 3258
 651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667
 3259 GAAAAATCCAAAGATT...AACTTAACCTGAT.....GCTTATTCACAAAT 3299
 667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684
 3300 CAAAGAAATCAAGGATCTCTAC..... 3321
 684 alaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
 3322GAGACTACTCTCAGTCTCTTTCGCAAGCAAGCAATTCATAATTTGGAT 3366


```

19  VallysGluIleSerAspLysIleSerLysGluMetValArgLeu 35
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1363 ATCAAGGTCATCCGACTCCCTAAATCCGAAAGCAAGAAATTTCTGC 1412
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35  uLysMetValLys...ThrPheMetAspMet...AspGlnAsps 49
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1413 AGAATGTCCTAAAGCAAGCTTTAGTGGACCTTTAGAAATCACAATGA 1462
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49  erGluGluLysGluLeu...TyrLeuAsnLeuAlaLeu 61
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1463 ACGCTGTAAAAGAAGAATTGAACAGTATAAGAGAGAGTTTGAACACTGCAAG 1512
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62  HisLeuAlaSerAspPheLeuLysHisProGly... 73
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1513 GTAATAGCAGATGATTCAAAAAACAACCTCCCGAAAATGAAGACTTACT 1562
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
74  LysAspValArgLeuLeu...ValAlaCysCysLeuAlaAspI 87
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1563 TAAGGAATTGCAGTTACGAAAGAAAAATTAGCACATCGCAAAAGAAAT 1612
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
87  lePheArgIle...TyrAlaProGluAlaProTyrThr 98
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1613 GTCACGTTTATCCTCTATAACTCACGAAGCAGATGAAGATATGAATAAT 1662
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
99  SerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLys 115
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1663 TTATCTGCAAAATCTAGTCTGTATTTATATCTCTGAAGAAACAAATTAAT 1712
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
115  sGlyLeuGluAspThrLys...SerProGlnPheAsnArgTyrP 129
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1713 TAAAGAAAGGCCTACCAAGGAACATCTTCAAAATCAAAATTTGAACATTCA 1762
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129  heTyrLeuLeuGluAsn...IleAlaTrpValLysSerTyrAsnIleCys 144
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

291 lnValValLysLeuLeuAlaLysMetPheClyAlaLysAspSerGlu... 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3708 ACTTCTCTCCATCTCTTAAATAATGGTAGTAAGCAAGGTGAAGAA 3757
307 .....LeuAlaSerGI 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3758 GATTACAGACCAGCTGTTGAAATTTTGTTCAGACCTTCTTTCAACAGT 3807
310 nAsnLysProLeuTyrGlnCys.....TyrLeuGlyA 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3808 CANTAAGCCTGAATGCCAGCTCTCAACTACTCTTACTCTTGTAGGGA 3857
321 rgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAla 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3858 GA..... 3859
338 SerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTy 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3860 .....CTGTGTGTTTCATGCTAGTTCAGTAACAAGTCAACAGAGATGGC 3900
354 rLeuLysValArgSerHisAsp.....ProGluGluAlaIleA 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3901 TTTAAGAGTGGCATCTCTTCATTACCTTGAACTGTGCTGCACGCGCTAA 3950
367 rgHisAspValIleValSer..... 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3951 GAAAGATGCTGTACAGGCAAAATGGATCAAGGATCTATAGAACGCATT 4000
374 .....IleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHI 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4001 TTAAACAGGTTTCAGGAGGGAAGATGAATCCACAACATTCACAAAAGC 4050
388 sLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTyrArgVal. 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4051 ATTGCTTGATTACTTGATGAAACACACTGAGACTGATCTTCACACTAGTGT 4100
405 .....ArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4101 TTTCCTCGTAAA.....TTCTATATAGCCCGAGTGTTCGAGACACA 4141
420 AlaLeu.....GlnSerAl 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4142 ACTCTGGAACACAAAGCAATGAATCACAATAAGATGAAGAATCAATC 4191
424 aAlaGlyLysAspAlaAlaLysGlnIle..... 433
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4192 TGAAGCAACACATCATGCAAGGAAATTTGAGACAACCTGGCCAAATTTATGC 4241
434 .....AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSer 448
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4242 ATCGAGCTGAAACACCGAAAAAGTTTCTTAGAAGCATTTATCAAAACCCA 4291
449 .....IleAs 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4292 CCTTCTCAGTTTAGCACATTAAGATGAACCTGATGACTGTGGACTATGA 4341
450 pAspArgLeuLeuValGluArgIle.....PheAlaG 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4342 TGATGCTGTGTTGATTGTTTCGATACTTGGCCCTCCATGAGGCCGTTTGCCC 4391
461 lnTyrMetValProHisAsnLeuGlu..... 469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4392 AGAGCTTTGATATTTATTTGACACAGATCCTACGAGTTCTTGTGTGAAAT 4441
470 .....ThrThrGluArgMetLysCysLeuTyr..... 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4442 GCAAATTCGTTCGACACAAAGCCATGAAGTGTTCGTGAGGTGTTGTC 4491
479 .....TyrLeuTyrAlaThrLeuAspLeuAsn..... 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4492 TGTAGACCCAGATTTCTAGCAAGGCTTGATATGCAACGAGGTGTTTCATG 4541

```

```

488 .....AlaValLysAla 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4542 GAGCATTTGATGGATAATTCGACTAGTGTCCGAGACGACGAGTAGAATTA 4591
492 LeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAs 508
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4592 CTAGGTCGATTTGCTTGTTCGACCTCAGCTTGTCTGAACAGATATTATGA 4641
508 pLeuLeuAspLeuIleLeuGlnProLysThrAspAlaSerValLysAlaI 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4642 TATGCTG...ATTGAAAGAATATTGGATACTGTGTATTCAGTGTCAAG... 4684
525 lePheSerLysValMetValIleThrArgAsnLeu.....Pro 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4685 .....AAAAGAGTAATAAGATTTCTCAGACATTTGTATTGAACAACCA 4729
538 AspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGI 554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4730 ACATTTCCAAAAATCACAAAATGTGTGTAATAATGATTCGCAGAGTCAA 4779
554 uAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThrC 571
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4780 TGATGAAGAGGCGATTAAAGAA..... 4801
571 ysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeu 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4802 .....TTAGTAATAAGAAACATTTCCAGAAACTC 4828
588 GlyAsnProLysGlnProThrAsnPro..... 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4829 TGGTTTACTCCAACTCCACAATGACAAAGAAGCAATGACAAGGAAAT 4878
596 ..... 596
4879 TTTAAACATTACCGATGTGGTTGCACATGCACAGATACTGGATATGACT 4928
597 ...PheLeuGluMetIleLysPheLeuLeuGlu.....Arg 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4929 GGTTCGAGCAACTGCTTCAAAACTTGTGAAAGTCTGAAGAGAGATTCTCTCA 4978
608 IleAlaProVal...HisIleAspThrGluSerIleSerAlaLeuIleLy 623
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4979 TATAAACCCTGTGAAGAAAGCTTGTACTCAACTTGTGATAACCTAGTTGA 5028
623 sGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyV 640
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5029 GCACATCTTAAATATGAGGAATCTCTAGCTGACTCTGACAATAAAGGTG 5078
640 aProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeu 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5079 TGAATTTCTGGAAGATTGGTA...GTTGCATAACCACTTTGTTCTTATTC 5125
657 SerPheThrHisPro...IleSerPheHisSerAlaGluThrPheGluSe 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5126 AGCAAAATAAGACCCAGCTCATGGTTAAACATGCAATGACTATGCAACC 5175
672 rLeuLeuAla.....CysLeuLysMetAspAsp..... 681
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5176 ATACCTTACCACTTAAATGTAGTAGCAGCAAAATGATTTTCATGTTATCTGCA 5225
682 .....GluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5226 ATGTTCAAAAATTCCTAGAGCTAGTTGTACCACCTGATGGAGCATCCAAGT 5275
697 .....SerLysIleGluGluAspPheProHis.....IleAr 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5276 GAAACTTTTCTGCCACTATTGAGGAGAGATCTAATGAAGCTCATCATCAA 5325
707 gSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProA 724
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5326 ATATGGCATGACTGTAGTGCACAT.....5350
724 rgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740

```


5995 A.....AATGTCATTG 6005
923 laGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyr 939
|||.....|
6006 CATTGACTCTAAATCAAGCTCTATTTCAT.....CCAGTTCAGTCT 6046
940 MetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
|||.....|
6047 GTGCCATATTAAATTCGTATGGCCACAGACCCAGAACCTGCTATGCGGAA 6096
956 aHisAlaArgGlnCysLeuValLysAsnIleAsnValArgGluTyrL 973
|||.....|
6097 CAAGGCTGATCAGCACTTGTGAATAGACAAATAATATGCTGGATTCA 6146
973 euLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuProGlu 989
|||.....|
6147 TTCATATGAAGCAGTGGCTGTATGAAGATG.....TCT 6181
990 TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa 1006
|||.....|
6182 TACCAGGTACAACAGCAATCAACACATCCCTAAAGATCCT..... 6223
1006 lLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp 1023
|||.....|
6224GTAAAGGGTTTCAGACAA...CACAGTCCCTCAGCGCTTGTGT 6266
1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPhe 1039
|||.....|
6267 CACACCTTTACTCCATCCAGTCCGTTGGAAACCGCCACACAGACGAGCCTTT 6316
1040 IleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPr 1056
|||.....|
6317 CTTATTCTTCTTACTCAACCTCTTT.....GA 6342
1056 oAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaLam 1073
|||.....|
6343 TGACACAGCAAA.....ACAGAGCTGACTA 6368
1073 etAsnIle...IleMetSerLysSerThrThrTyrSerLeuGluSerPro 1088
|||.....|
6369 TGCTCTGTATATAGCAGACAATCTAGCCTGTTTCCATACCAGACACAG 6418
1089 LysAspProValLeuProAlaArgPheThrGlnProAspLysAsnPh 1105
|||.....|
6419 GAAGAGCCGTGTTT..... 6433
1105 eSerAsnThrLysAsnTyrLeuProGluMetLysSerPheThrP 1122
|||.....| 6433
1122 roGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeu 1138
|||.....|
6434ATAATGCATCATATAGACATTACACTC 6460
1139 SerSerAlaGly.....LysGlnSerGlnThrLy 1148
|||.....|
6461 TCAGTTCCTGGTAGTAACCTACTGTCAGTCATTCAAGGAGTCTATGTTAA 6510
1148 sSerSerArgMetGluThrValSerAsnAlaSer..... 1159
|||.....|
6511 GGCAAAAGGAAGAGAGAAATCATCCTAGTAGGAATGAGTCAA 6560
1160SerSerAsnProSerSerProGlyArgIleLysGlyArgLeu 1174
|||.....|
6561 GCGACAGTGAAGAAGATTTCAGGCCCTCGGAAGTCACGGAACCTGTA 6610
1175 AspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSe 1191
|||.....|
6611 GATTTCAGATTCAGATTCAAGTACAGAGACGATATAAATTCAGTGTGAA 6660
1191 rProLeuProGluLysLysSerAspLysArgAspSerAspLeuValA 1208
|||.....|
6661 ATGTTTCCAGAAAATTCAGTCTCTTATTCGNAATTTGCAATGTC.... 6706

1208 rgSerGluLeuGluLysProArgGlyArgLysLysThrProValThrGlu 1224
|||.....|
6707TCCCGAGGTATTATTACTTCTCATGTTAAAA 6739
1225 GlnGluGluLys.....LeuGlyMetAspAspLeuThrLysLeuValG 1239
|||.....|
6740 CAACATTGAAGAATCTTGTGGATTTCCTGAT.....AGTAAATTC 6783
1239 nGluGlnLysProLysGlySerGlnArg.....SerA 1250
|||.....|
6784 GAAGTACTCTCCATCTGAATCTGCAAAAGTATATGATAAGCGATAACC 6833
1250 rgLysArgGly..... 1253
6834 GAAACACAGGAGTCTCATTTTTCATCCAAAAACACACTGGACTTCCTCGCG 6883
1254HisThrAlaSerGluSerAspG 1261
6884 AGTGACATGGCTAATTCCAAATACAGAGAGGTGAAAAGGAGTATAGT 6933
1261 uGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnG 1278
|||.....|
6934 AAACAGTATCTAGATTTCAAACTTCTCATGACATCTGGACCTGATG 6983
1278 luAspGluGlnAsnSerProProLysLysLysArgGlyArgPro... 1293
|||.....|
6984 AAGAAGAAGAAGGGAGGTTTCACTAGTAGCACAATGCTCGGAACAAA 7033
1294ProLysProLeuGlyGlyThrProLysGluGluProThrMe 1308
|||.....|
7034 GCAATTACCTCACTGCTGGAGGAGGCAGCCCTAAATAATACACAGC 7083
1308 tLysThrSerLysLysGlySerLysLysSerGlyProProAlaProG 1325
|||.....|
7084 AGAGACAGAA.....G 7094
1325 luGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSer 1341
|||.....|
7095 ATGATGAAAGTATGGGAGGATAGAGAGGAGGAGGACTTCAGGGGTGAGG 7144
1342 LysSerLysGlnHisArgValSerArgArg 1351
|||.....|
7145 CGAGAGGAGGATCAAGCTATTTCGACGCGT 7174
seq_name: gb_prl:AB019494
seq_documentation_block:
LOCUS AB019494 7821 bp mRNA partial cds. PRI 17-JUN-1999
DEFINITION Homo sapiens IDN3 mRNA, partial cds.
ACCESSION AB019494
VERSION AB019494.1 GI:4760548
KEYWORDS IDN3.
SOURCE Homo sapiens adult male testis cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Aihara,T., Yasuo,M., Kumiko,K., Sasaki,Y., Imaoka,S., Monden,M. and Nakamura,Y.
TITLE Genes preferentially expressed in precancerous lesion of HCC
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 7821)
AUTHORS Aihara,T., Miyoshi,Y. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) to the DDBJ/EMBL/GenBank databases. Yusuke Nakamura, Institute of Medical Science, University of Tokyo; 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan (E-mail:yusuke@ims.u-tokyo.ac.jp, Tel:81-35449-5372, Fax:81-35449-5433)
FEATURES
Location/Qualifiers
1..7821
/organism="Homo sapiens"

misc_feature 7521..9277
/note="assembly_fragment"
misc_feature 9378..11138
/note="assembly_fragment"
misc_feature 11239..13265
/note="assembly_fragment"
misc_feature 13366..15440
/note="assembly_fragment"
misc_feature 15541..17308
/note="assembly_fragment"
misc_feature 17409..19894
/note="assembly_fragment"
misc_feature 19995..21852
/note="assembly_fragment"
misc_feature 21953..24275
/note="assembly_fragment"
misc_feature 24376..27231
/note="assembly_fragment"
misc_feature 27332..31148
/note="assembly_fragment"
misc_feature 31249..33450
/note="assembly_fragment"
misc_feature 33551..36754
/note="assembly_fragment"
misc_feature 36855..39988
/note="assembly_fragment"
misc_feature 40089..42367
/note="assembly_fragment"
misc_feature 42468..45775
/note="assembly_fragment"
misc_feature 45876..49987
/note="assembly_fragment"
misc_feature 50088..53225
/note="assembly_fragment"
misc_feature 53326..56622
/note="assembly_fragment"
misc_feature 56723..59393
/note="assembly_fragment"

alignment_scores:
Quality: 314.00 Length: 77
Ratio: 4.486 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 80.519
alignment_block:
US-09-512-581-2 x AC022862 ..

Align seg 1/1 to: AC022862 from: 1 to: 160962

35 LeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGluG1 51
||||:|||||
10880 TTACAGATGGTAGTGAACACCTTTATGATATGATCAGGACTCAGAGA 10929
||||:|||||
10930 TGAACACAGCAGATATCCACACTAGCCTTGATCTGATCTGAATCT 10979
||||:|||||
68 heLeuLysHisProGlyLysAspValargLeuValAlaCysCysLeu 84
||||:|||||
10980 TCCCTCAGAACCCCAATAAAGATGCTGCTCCTTGATGATGTTG 11029
||||:|||||
85 AlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAs 101
||||:|||||
11030 GCTGATATCTTCGTATCTATGCCCCAGAGACTCCATATACTCCCATGA 11079
||||:|||||
101 pLysLeuLysAspIlePheMetPheIleThr 111
||||:|||||
11080 TAAACTTAAGCTAAATACAGCCTTTTCACT 11110
||||:|||||

seq_name: gb_p11-AF049529

seq_documentation_block:
LOCUS AF049529 1431 bp mRNA PLN 05-JAN-1999

DEFINITION Schizosaccharomyces pombe unknown mRNA, partial cds.
ACCESSION AF049529
VERSION AF049529.1 GI:4105678
KEYWORDS
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe

REFERENCE 1 (bases 1 to 1431)
AUTHORS Lee,M., Yoo,H.S. and Chung,K.S.
TITLE clone 17 (bimD homologue)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1431)
AUTHORS Lee,M., Yoo,H.S. and Chung,K.S.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-1998) Cell cycle & Signal Research Unit, Korea Research Institute of Bioscience and Biotechnology, P.O. Box 115, Yuseong, Taejeon 305-600, Republic of Korea

FEATURES
Source
1..1431
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
/clone="17"
CDS
<1..1173
/note="similar to bimD protein"
/codon_start=1
/product="unknown"
/protein_id="AAD02493.1"
/db_xref="GI:4105679"
translation="GTSKVLAIKLVNRLRAAAGTEALNIGAIKILKLVLMADGE
LSPFKNTPKISRAYLRLTASKYPFLKICSIPIFAEHIDFSYVSNLLIMPDENFDVRL
FLTKLOLQOLKLPISYVPLFLFADVPEELKTKASINIRSOVAFOKTHDFTWEY
VATYLIHLSSHDPDISSENSLDRIATIRYVDFVNVSENVPIVFLMQRKQSYD
VIEDGNNYIYVLSDMAQKILQVKQNFQWSLTYPKQILPEYLILPISIDKKRIF
NKFITPKMESQIEHAIRTPVSSFAKQTTNKHANLKQKTHSKSKDKKSSRRKNEKR
RKLNEQNPINRVPERSSRFQIRINYSAPSSSEISEEISEEIDFDEIDL"

BASE COUNT 503 a 240 c 254 g 434 t
ORIGIN

alignment_scores:
Quality: 309.50 Length: 424
Ratio: 1.253 Gaps: 11
Percent Similarity: 58.255 Percent Identity: 24.764

alignment_block:
US-09-512-581-2 x AF049529 ..

Align seg 1/1 to: AF049529 from: 1 to: 1431

825 LysIleGlnAlaIleLysMetMetValargTrpLeuLeuGlyMetLysAs 841
||||:|||||
10 AAAGTTTATGCAATAGGTTTGGTTAATCGACTAGGCGCGGCAGG 59
||||:|||||
841 nAsnHisSer.....LysSerGlyThrSerThrLeuArgLeuLeuThr 856
||||:|||||
60 AGGAACCAACCGTAAATATTGGAGCTCCTATCATTAAGCTATTGAAAG 109
||||:|||||
856 hrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
||||:|||||
110 TATTGTTAATGGCTGACGCTGAATTAAGTCGCTCAACAATACTCCCAA 159
||||:|||||
873 ProAspMetSerArgLeuArgLeuAlaIleGlySerAlaIleValLysLe 889
||||:|||||
160 ATTTCTCGTGTCTTTATACGCTCTCAGCTAGCAAAATATTTTAAAGTT 209
||||:|||||
889 uAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnL 906
||||:|||||
210 ATGTTCCATTCCTTTTATGCGGACATATAGATTTTAGTCTTACGTTT 259
||||:|||||
906 euCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPhe 922
||||:|||||

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 DeRellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., Olivar, T.M., Peterson, K.,
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.

Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced gi:6978239.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6193
 Center clone name: 343_C_9
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 145534 bases at least Q40
 Consensus quality: 151792 bases at least Q30
 Consensus quality: 154567 bases at least Q20
 Insert size: 194000; agarose-fp
 Insert size: 157262; sum-of-contigs
 Quality coverage: 3.2 in Q20 bases; agarose-fp
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

TITLE JOURNAL

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 38 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 1187: contig of 1187 bp in length
* 1188 1287: gap of 100 bp
* 1288 2962: contig of 1675 bp in length
* 2963 3062: gap of 100 bp
* 3063 4458: contig of 1396 bp in length
* 4459 4558: gap of 100 bp
* 4559 5583: contig of 1025 bp in length
* 5584 5683: gap of 100 bp
* 5684 7420: contig of 1737 bp in length
* 7421 7520: gap of 100 bp
* 7521 9277: contig of 1757 bp in length
* 9278 9377: gap of 100 bp
* 9378 11138: contig of 1761 bp in length
* 11139 11238: gap of 100 bp
* 11239 13265: contig of 2027 bp in length
* 13266 13365: gap of 100 bp
* 13366 15440: contig of 2075 bp in length
* 15441 15540: gap of 100 bp
* 15541 17308: contig of 1768 bp in length
* 17309 17408: gap of 100 bp
* 17409 19894: contig of 2486 bp in length
* 19895 19994: gap of 100 bp

```

```

* 19995 21852: contig of 1858 bp in length
* 21853 21952: gap of 100 bp
* 21953 24275: contig of 2323 bp in length
* 24276 24375: gap of 100 bp
* 24376 27231: contig of 2856 bp in length
* 27232 27331: gap of 100 bp
* 27332 31148: contig of 3817 bp in length
* 31149 31248: gap of 100 bp
* 31249 33450: contig of 2202 bp in length
* 33451 33550: gap of 100 bp
* 33551 36754: contig of 3204 bp in length
* 36755 36854: gap of 100 bp
* 36855 39988: contig of 3134 bp in length
* 39989 40088: gap of 100 bp
* 40089 42367: contig of 2279 bp in length
* 42368 42467: gap of 100 bp
* 42468 45775: contig of 3308 bp in length
* 45776 45875: gap of 100 bp
* 45876 49987: contig of 4112 bp in length
* 49988 50087: gap of 100 bp
* 50088 53225: contig of 3138 bp in length
* 53226 53325: gap of 100 bp
* 53326 56622: contig of 3297 bp in length
* 56623 56722: gap of 100 bp
* 56723 59393: contig of 2671 bp in length
* 59394 59493: gap of 100 bp
* 59494 63369: contig of 3876 bp in length
* 63370 63469: gap of 100 bp
* 63470 68148: contig of 4679 bp in length
* 68149 68248: gap of 100 bp
* 68249 72051: contig of 3803 bp in length
* 72052 72151: gap of 100 bp
* 72152 78079: contig of 5928 bp in length
* 78080 78179: gap of 100 bp
* 78180 82389: contig of 4210 bp in length
* 82390 82489: gap of 100 bp
* 82490 87187: contig of 4698 bp in length
* 87188 87287: gap of 100 bp
* 87288 93791: contig of 6504 bp in length
* 93792 93891: gap of 100 bp
* 93892 99177: contig of 5286 bp in length
* 99178 99277: gap of 100 bp
* 99278 104756: contig of 5479 bp in length
* 104757 104856: gap of 100 bp
* 104857 112644: contig of 7788 bp in length
* 112645 112744: gap of 100 bp
* 112745 119589: contig of 6845 bp in length
* 119590 119689: gap of 100 bp
* 119690 130279: contig of 10590 bp in length
* 130280 130379: gap of 100 bp
* 130380 144451: contig of 14072 bp in length
* 144452 144551: gap of 100 bp
* 144552 160962: contig of 16411 bp in length.

```

FEATURES source

```

1..160962
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-343C9"
/clone_lib="RPC1-11 Human Male BAC"
1..1187
/note="assembly_fragment"
1288..2962
/note="assembly_fragment"
3063..4458
/note="assembly_fragment"
4559..5583
/note="assembly_fragment"
clone_end:T7
vector_side:right
5684..7420
/note="assembly_fragment"

```

```

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```


Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
 Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10823

Center clone name: 2504_H2

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 135676 bases at least Q40
 Consensus quality: 149264 bases at least Q30
 Consensus quality: 155372 bases at least Q20
 Insert size: 219000; agarose-fp
 Insert size: 159671; sum-of-contigs
 Quality coverage: 2.3 in Q20 bases; agarose-fp
 Quality coverage: 3.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 consists of 81 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 162: contig of 162 bp in length
 163 262: gap of 100 bp
 263 674: contig of 412 bp in length
 675 774: gap of 100 bp
 775 1187: contig of 413 bp in length
 1188 1287: gap of 100 bp
 1288 1965: contig of 678 bp in length
 1966 2065: gap of 100 bp
 2066 2205: contig of 140 bp in length
 2206 2305: gap of 100 bp
 2306 2596: contig of 291 bp in length
 2597 2696: gap of 100 bp
 2697 3096: contig of 400 bp in length
 3097 3196: gap of 100 bp
 3197 3778: contig of 582 bp in length
 3779 3878: gap of 100 bp
 3879 4519: contig of 641 bp in length
 4520 4619: gap of 100 bp
 4620 5136: contig of 517 bp in length
 5137 5236: gap of 100 bp
 5237 5877: contig of 641 bp in length
 5878 5977: gap of 100 bp
 5978 6662: contig of 685 bp in length
 6663 6762: gap of 100 bp
 6763 7437: contig of 675 bp in length
 7438 7537: gap of 100 bp
 7538 8181: contig of 644 bp in length
 8182 8281: gap of 100 bp
 8282 8720: contig of 439 bp in length

8721 8820: gap of 100 bp
 8821 9297: contig of 477 bp in length
 9298 9397: gap of 100 bp
 9398 10235: contig of 838 bp in length
 10236 10335: gap of 100 bp
 10336 10985: contig of 650 bp in length
 10986 11085: gap of 100 bp
 11086 12119: contig of 1034 bp in length
 12120 12219: gap of 100 bp
 12220 13370: contig of 1151 bp in length
 13371 13470: gap of 100 bp
 13471 14862: contig of 1392 bp in length
 14863 14962: gap of 100 bp
 14963 15547: contig of 585 bp in length
 15548 15647: gap of 100 bp
 15648 16772: contig of 1125 bp in length
 16773 16872: gap of 100 bp
 16873 18277: contig of 1405 bp in length
 18278 18377: gap of 100 bp
 18378 19739: contig of 1362 bp in length
 19740 19839: gap of 100 bp
 19840 21158: contig of 1319 bp in length
 21159 21258: gap of 100 bp
 21259 22237: contig of 979 bp in length
 22238 22337: gap of 100 bp
 22338 23288: contig of 951 bp in length
 23289 23388: gap of 100 bp
 23389 24797: contig of 1409 bp in length
 24798 24897: gap of 100 bp
 24898 25704: contig of 807 bp in length
 25705 25804: gap of 100 bp
 25805 27367: contig of 1563 bp in length
 27368 27467: gap of 100 bp
 27468 28651: contig of 1184 bp in length
 28652 28751: gap of 100 bp
 28752 30113: contig of 1362 bp in length
 30114 30213: gap of 100 bp
 30214 31445: contig of 1232 bp in length
 31446 31545: gap of 100 bp
 31546 32766: contig of 1221 bp in length
 32767 32866: gap of 100 bp
 32867 34495: contig of 1629 bp in length
 34496 34595: gap of 100 bp
 34596 36032: contig of 1437 bp in length
 36033 36132: gap of 100 bp
 36133 37465: contig of 1333 bp in length
 37466 37565: gap of 100 bp
 37566 38460: contig of 895 bp in length
 38461 38560: gap of 100 bp
 38561 39981: contig of 1421 bp in length
 39982 40081: gap of 100 bp
 40082 41219: contig of 1138 bp in length
 41220 41319: gap of 100 bp
 41320 43417: contig of 2098 bp in length
 43418 43517: gap of 100 bp
 43518 44913: contig of 1396 bp in length
 44914 45013: gap of 100 bp
 45014 46461: contig of 1448 bp in length
 46462 46561: gap of 100 bp
 46562 47955: contig of 1394 bp in length
 47956 48055: gap of 100 bp
 48056 49780: contig of 1725 bp in length
 49781 49880: gap of 100 bp
 49881 51344: contig of 1464 bp in length
 51345 51444: gap of 100 bp
 51445 53900: contig of 2456 bp in length
 53901 54000: gap of 100 bp
 54001 55543: contig of 1543 bp in length
 55544 55643: gap of 100 bp
 55644 57360: contig of 1717 bp in length
 57361 57460: gap of 100 bp
 57461 59760: contig of 2300 bp in length
 59761 59860: gap of 100 bp

63063 AAGCCAGATAGCTATGTTCCAGGGATAAATAGCAGATTAAATGATACTTT 63112

Marquis, N., Matthews, C., McCa
McPheeters, R., Meldrum, J., Me

Leibovitz, J., Dev
Marquis, N., Matthe


```

misc_feature      15116..16652
/note="assembly_fragment"
misc_feature      16753..19184
/note="assembly_fragment"
misc_feature      19285..21169
/note="assembly_fragment"
misc_feature      21270..24243
/note="assembly_fragment"
misc_feature      24344..27226
/note="assembly_fragment"
misc_feature      27327..30105
/note="assembly_fragment"
misc_feature      30206..32466
/note="assembly_fragment"
misc_feature      32567..35504
/note="assembly_fragment"
misc_feature      35605..38079
/note="assembly_fragment"
misc_feature      38180..41205
/note="assembly_fragment"

alignment_scores:
  Quality: 353.50      Length: 302
  Ratio: 3.243        Gaps: 3
  Percent Similarity: 36.093      Percent Identity: 29.139

alignment_block:
US-09-512-581-2 x AC023006 ..

Align seg 1/1 to: AC023006 from: 1 to: 162775

913 AspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysG1 929
|||||
143786 GATGAGTGTACCAAGTACGAGTATGTCGAGAGCTGCATAAGC 143835
|||
929 YLeuSerArgLeuArgLeuProLeuGluTyrMetAlaileCysAlaLeuC 946
|||
143836 ACTGTGTAAGTACTGTCCTCCATGGAGTATATGCGCATCTTTGCCCTGT 143885
|||
946 YSAlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeu 962
|||
143886 GTGCCAAGATCCTGTGAAGGAGAGACACACGACGACATGTTTA 143935
|||
963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaLa.V 979
|||
143936 CTGAAATATACGATATACGACGAGGAATACATTAAGCAGATCTATGCG 143985
|||
979 al..... 979
:
143986 TACTGTGAAGTAACTCAGAGTTCCTCGTGTCGAAATATATTTCTTAATG 144035
|||
979 ..... 979

144036 TATCCATTTTATAGTGTGTTTTCCTGTCATGTTTCTATTATAGTTT 144085
|||
979 ..... 979

144086 ATATAAACTGGAGGTTTATAGAATGCTGTCATCTAGTATTATGATGTA 144135
|||
979 ..... 979

144136 CTTAAAAAAGTAACTCCTGAGTCTCTATTTATGTCGATTAACCTGCGCAA 144185
|||
979 ..... 979

144186 ACTGCATTTTTCCTATGGATCAGGAGTGTAGTTTAATTTGAACATATTG 144235
|||
979 ..... 979

144236 CTCGGATGTCAGACTAAATAATTTTTCCTAGCTTTCGACGGGACCTGTT 144285
|||
979 ..... 979

```

```

144286 GTTAAATGGGTTTAAATATCTGATACCTTTTGACITTCAGTATCTTTCAAAG 144335
979 ..... 979

144336 TGAATACAACTGCATACCAAGTACCAGAGAAAAAGAAATGAAAGTGAAG 144385
979 ..... 979

144386 TGTAAAGTATTTTAGTATCCAGGTATCATGTTGTGCCCTAATTTAGAATA 144435
979 ..... 979

144436 ACTGAGGAAAAAATGAAACTTAAGTGGTTAGAACTCAGTCCTCACTAAGG 144485
980 .....SerGluLysLeuLe 984
144486 CATTTTATAAAGGTGTTTAAATAATGTTTCTTACCTTCAGAGAAATATT 144535
984 uSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
144536 ATCAGTGTGCTGGAATATGTTAGTTCATACATGATTCACCTGCTAGCCC 144585
1001 isAspProAspTyrValLysValcInAspIleGluInLeuLysAspVal 1017
144586 ATGATCCAGATTTTACAAGATCAAGATGTTGATCAGCTTCGTGATATC 144635
1018 LysGluCysLeuTrpPheValLeuGluLeuMetAlaLysAsnGluAs 1034
144636 AAAGAGTAAGTC.....TTGTTGTTTTTCAAGAAGACTCTAA 144670
1034 nAsn 1035
144671 AAAT 144674

seq_name: gb_htg14:AC068352
seq_documentation_block:
LOCUS AC068352 145450 bp DNA HTG 25-JUN-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-168F19 map 4, WORKING DRAFT
SEQUENCE, 31 unordered pieces.
ACCESSION AC068352
VERSION AC068352.2 GI:8705054
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145450)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-168F19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145450)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bozulavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Gallagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lechoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```


COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6997326.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L6608
Center clone name: 815_B_22
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 138867 bases at least Q40
Consensus quality: 151127 bases at least Q30
Consensus quality: 155740 bases at least Q20
Insert size: 194000; agarose-fp
Quality coverage: 2.8 in Q20 bases; agarose-fp
Quality coverage: 3.4 in Q20 bases; sum-of-contigs

*** NOTE: This is a 'working draft' sequence. It currently
consists of 40 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be revised.

| | | |
|-------|--------|-----------------------------|
| 1 | 1131: | Contig of 1131 bp in length |
| 1132 | 1231: | gap of 100 bp |
| 1232 | 2342: | Contig of 1111 bp in length |
| 2343 | 2442: | gap of 100 bp |
| 2443 | 3544: | Contig of 1102 bp in length |
| 3545 | 3644: | gap of 100 bp |
| 3645 | 4649: | Contig of 1005 bp in length |
| 4650 | 4749: | gap of 100 bp |
| 4750 | 5804: | Contig of 1055 bp in length |
| 5805 | 5904: | gap of 100 bp |
| 5905 | 7039: | Contig of 1135 bp in length |
| 7040 | 7139: | gap of 100 bp |
| 7140 | 7359: | Contig of 220 bp in length |
| 7360 | 7459: | gap of 100 bp |
| 7460 | 9365: | Contig of 1906 bp in length |
| 9366 | 9465: | gap of 100 bp |
| 9466 | 10617: | Contig of 1152 bp in length |
| 10618 | 10717: | gap of 100 bp |
| 10718 | 12064: | Contig of 1347 bp in length |
| 12065 | 12164: | gap of 100 bp |
| 12165 | 13096: | Contig of 932 bp in length |
| 13097 | 13196: | gap of 100 bp |
| 13197 | 15015: | Contig of 1819 bp in length |
| 15016 | 15115: | gap of 100 bp |
| 15116 | 16652: | Contig of 1537 bp in length |
| 16653 | 16752: | gap of 100 bp |
| 16753 | 19184: | Contig of 2432 bp in length |
| 19185 | 19284: | gap of 100 bp |
| 19285 | 21169: | Contig of 1885 bp in length |
| 21170 | 21269: | gap of 100 bp |
| 21270 | 24243: | Contig of 2974 bp in length |
| 24244 | 24343: | gap of 100 bp |
| 24344 | 27226: | Contig of 2883 bp in length |
| 27227 | 27326: | gap of 100 bp |
| 27327 | 30105: | Contig of 2779 bp in length |
| 30106 | 30205: | gap of 100 bp |
| 30206 | 32466: | Contig of 2261 bp in length |
| 32467 | 32566: | gap of 100 bp |
| 32567 | 33504: | Contig of 2938 bp in length |
| 35505 | 35604: | gap of 100 bp |

198 ThrValLeu.....ValAsnLeuValProLahI 207
1475 TTAATTTTAAATAAGTTTGTGACCTATAACCCGAATCAAAATTCCTCAAGG 1426
207 slyAsnLeuAsnLysGln...AlaTyrAspLeuAlaLysAlaLeuLeuL 223
1425 TTTGAACGGTACTCTGACCTGGCTGACCAAGTTAGCTAAATTTTGTGTG 1376
223 ysArgThrAlaGlnAlaIleGluProTyrIleThrPhePheAsnGln 239
1375 ATACTTATCCATAGATGACGTAGACATCACTAAATACATCTATCCGAG 1326
240 ValLeuMetLeu.....GlyLysThrSerIleSerAspLe 251
1325 ATATTTCATGAACCAACTAATGATGACAAATAATCAAGACTTCTCATGT 1276
251 userGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerH 268
1275 GTGAGTGAATTAACATAACTGGTATGAGACTTTGGGAACAGTACCCG 1226
268 isLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsn 284
1225 AACTGATCAATGACGTAATCGGATTTATATACCACGAATATCTTCGGA 1176
285 AspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheG 301
1175 AAGGAGCTTTTCAGAAAGAAAGCTACAAAATTAATTTGGTCAAACTTAC 1126
301 yAla...LysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnC 317
1125 CTCATATTCGGATTTAAATTTGTTTTCACACATTCGATCCGATCCGAT 1076
317 ystyrLeuGluArgPheAsnAspIleHisValProIleArgLeuGluCys 333
1075 CATGGATTTCCAAATTCGTATATGATAGTCCCGATGTTAGATGAGATG 1026
334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys 350
1025 ACAGAATCAATTCCTCAATCATTTGCGACGAGAGAGACATATCCAAAGA 976
350 pLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 367
975 ACTGAATCAGGACTGCGCAAAACCTTTATGATCTGATCTCGTGATC 926
367 rgHisAspValIleValSerIle.....ValThrAlaAlaLys 379
925 GCAGGACTTCTGTATGATTTTAAATAAGTTCCCGTTACTGAAATTTGG 876
380 LysAspIleLeu.....LeuValAsnAspHisLeuLeuAsnPheValar 394
875 AAAAATATAACCAATAAAGCCATTTATACATCTTTACTTCTATTTGCCCG 826
394 gGluArgThrLeuAspArgTrpArgValArgLysGluAlaMetMetG 411
825 AGAA.....AAACATPAAAGAGGTAGGGAACCTCTGCAATTAACA 788
411 lyLeuAlaGlnIleTyrLysLysTyrAlaLeuGln.....SerAla 424
787 CTATGCCCAAGTTTACTCAAAATTCATTTGAATGAGATCGAAAGGACATAC 738
425 AlaGlyLysAspAlaLysGlnIleAlaTrpIleLysAspLysLeuLe 441
737 CAAATATAAGAAATTTGGGAAATCATPAGACACCATACCATCCACTTTATA 688
441 uHisIleTyrTrpGlnAspSerIleAspArgLeuLeuValGluArgI 458
687 CAATCTCTATTATCAATGATTTTAAACATCAATGAAAGTCCGACGCG 638
458 lePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMet 474
637 TAATTTTCAATACCTACTACCGTTTGAACCCAGATTAACGATAAAAGAGTC 588

475 LysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAl 491
587 CACAGACTGCTAAGCTTTGTGCACATTTTGAAGAAGGCATTCACTTC 538
491 aleuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysA 508
537 ATTTTTCGCTTTCAATGCAAGCAAAATCAAAATATCTTCGCTATATCCA 488
508 splLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAla 524
487 AATATATTGATTTTACTAAATTTTGAACAACATCAAGAAAGTATGAGCTCG 438
525 IlePheSerLysValMetVal.....I 532
437 TCTCAAGAGTCTTATGTAATGAACAAGTACAAACAACTCTCCAGTGGT 388
532 eThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysP 549
387 AGCTTCTGCGCTTCTGACTCTACGAAGCAATCGACGCCCTCGAAACAA 338
549 heThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluVal 565
337 TAAACAAATTT.....AAGCAGAAAGAAATCTTTTATTATTAATGCC 294
566 LeuValSerProThrCysSerCysLysGlnAlaGluGluCysValArgG 582
293 TGTGTCCACCAAGCATATTCGTTTCTTACTTTTAAAAAATGTTACAATGA 244
582 uIleThrLysLysLeuGlyAsnProLys..... 591
243 ATTGGTCAGTAAGTTTACAAACGCCAGGACTATTTAAAAAATATAATAT 194
592GlnProThrAsnProPheLeuGluMetIleLys 602
193 CTACTGCTGCTTCCATTAATGCCACGATGAT...ATCCCAAGGTAATACAA 147
603 PheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSe 619
146 ATTTGCTTTTGTAGAGCCTCACCATAATTTATATGATATCCCAACATCAG 97
619 rAlaLeuIleLysGlnValAsnLysSer 628
96 TGTGTTTATGAACTATCTAATAATTTCT 69

seq_name: gb_pr5:AK023592

seq_documentation_block: 1489 bp mRNA PRI 29-SEP-2000
LOCUS AK023592 Homo sapiens cDNA FLJ13530 fis, clone PLACE1006248, highly similar
DEFINITION to Homo sapiens mRNA for KIAA0648 protein.

ACCESSION AK023592
VERSION AK023592.1 GI:10435570
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
clone:PLACE1006248.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)
AUTHORS
Isoqai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K.,
Tanai, H., Kinata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y.,
Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahori, K.,
Masuho, Y. and Kanehori, K.

TITLE

NEDO human cDNA sequencing project
JOURNAL
REFERENCE
Unpublished (2000)
2 (bases 1 to 1489)
AUTHORS
Direct Submission
TITLE
Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isoqai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952).


```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA48113
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: pLasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159306 bases at least Q40
Consensus quality: 161143 bases at least Q30
Consensus quality: 162084 bases at least Q20
Insert size: 162842; sum-of-contigs
Insert size: 163676; 2.5% error; agarose-fp
Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality
coverage: 4.42x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 20261: contig of 20261 bp in length
* 20262 20361: gap of 100 bp
* 20362 38140: contig of 17779 bp in length
* 38141 38240: gap of 100 bp
* 38241 85018: contig of 46778 bp in length
* 85019 85118: gap of 100 bp
* 85119 93385: contig of 8267 bp in length
* 93386 93485: gap of 100 bp
* 93486 96470: contig of 2985 bp in length
* 96471 96570: gap of 100 bp
* 96571 137301: contig of 40731 bp in length
* 137302 137401: gap of 100 bp
* 137402 147035: contig of 9634 bp in length
* 147036 147135: gap of 100 bp
* 147136 149710: contig of 2575 bp in length
* 149711 149810: gap of 100 bp
* 149811 163642: contig of 13832 bp in length.
FEATURES
Source
Location/Qualifiers
1..163642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-448I13"
/clone_lib="RPC1-11.2"
1..20261
/note="assembly_fragment:00330
clone_end:T7
vector_side:left"
20362..38140
/note="assembly_fragment:00335
fragment_chain:1"
38241..85018
/note="assembly_fragment:01397
fragment_chain:1"
85119..93385
/note="assembly_fragment:01083
fragment_chain:1"
93486..96470
/note="assembly_fragment:00197
fragment_chain:1"
96571..137301
/note="assembly_fragment:00284
fragment_chain:2"
137402..147035
/note="assembly_fragment:00874
fragment_chain:2"

```

```

misc_feature 147136..149710
/note="assembly_fragment:00391
fragment_chain:3"
misc_feature 149811..163642
/note="assembly_fragment:00398
fragment_chain:3"
BASE COUNT 54756 a 31368 c 29374 g 47337 t 807 others
ORIGIN
alignment_scores:
Quality: 390.50 Length: 349
Ratio: 3.719 Gaps: 3
Percent Similarity: 30.086 Percent Identity: 29.799
alignment_block:
US-09-512-581-2 x AL353724/rev ..
Align seg 1/1 to reverse of: AL353724 from: 1 to: 163642
104 LysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspTh 120
|||||
26467 AAGGATATATTTATGTTTATACAGACAGTTGAAGGGGCTAGAGGATAC 26418
120 rLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu..... 133
|||||
26417 AAAGGCCACAAATCAATAGGTATTTTATTACTTGAGGTAAGCAATA 26368
133 ..... 133
26367 TCTGTATCTTGAGATGACATTTTAAACTGATTTTGTGTTGTATATT 26318
134 .....As 134
26317 TTAATAATCAAAATGCTTAACCTTTTACACCTTATTTTAGAA 26268
|||||
134 nileAlaTirpValLysSerTyrAsnIleCysPheGluLeuGluAspSerA 151
|||||
26267 CATGCTGGTCAAGTCATATACATATGCTTTGAGTTAGAGATAGCA 26218
151 snGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIle..Asn.. 166
|||||
26217 ATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAGTAA 26168
166 ..... 166
26167 GTTTATTTTAAAGTATGTAACATTTAAAAAAGGTAACAAATTTTGAAC 26118
166 ..... 166
26117 TTGTAGGTGAATCTGCATAAACTGTTACAATGAATATTAGAGGCTGAGA 26068
166 ..... 166
26067 GATGTAACCTTTTAACACACACTTTTATTTTATACATGTTTCAAAAATGAATA 26018
166 ..... 166
26017 GAGCTTAATTTTGTTATATATATGTAATATTGTTAGAGGAGTGCCATAAG 25968
166 ..... 166
25967 GCCAGTGTAAATTTTGTGATTCATTTTGTCTATATAAACATACAT 25918
166 ..... 166
25917 ACTGACTTCTTAATGCATATGGCTATTTGAGCTTTTGTGTATGCTTTAA 25868
166 ..... 166
25867 AATTATTTACTGGAATTAACCTTCAGAGCAATTCATTTTCTGAATTTAAG 25818
166 ..... 166

```



```

repeat_region /note="Alu repeat: matches 1. .308 of consensus"
75593. .75883
/partial
repeat_region /note="Alu repeat: matches 308. .1 of consensus"
75974. .76251
/partial
repeat_region /note="Alu repeat: matches 294. .1 of consensus"
76551. .76608
/note="MIR element fragment"
78482. .78832
/note="MER11B element fragment"
78574. .78743
/note="MER11A element fragment"
78941. .79021
/note="MER11B element fragment"
79249. .79709
/note="MER11A element fragment"
80233. .80302
/note="SVA element fragment"
80355. .80535
/partial
repeat_region /note="Alu repeat: matches 213. .20 of consensus"
80563. .80619
/partial
repeat_region /note="Alu repeat: matches 277. .221 of consensus"
80651. .82294
/note="SVA element fragment"
82753. .83043
/partial
repeat_region /note="Alu repeat: matches 308. .1 of consensus"
83149. .83296

```

alignment_scores:

```

Quality: 390.50 Length: 349
Ratio: 3.719 Gaps: 3
Percent Similarity: 30.086 Percent Identity: 29.799

```

alignment_block:

```

US-09-512-581-2 x HS267P19 ..

```

```

Align seg 1/1 to: HS267P19 from: 1 to: 113704

```

```

104 LysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspTh 120
|||||
22830 AAGGATATATTTATGCTTTATACAGACAGTTCGAAGGGCTAGAGGATAC 22879
120 rLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu..... 133
22880 AAGAGCCCAACATTCATAGATATTTTATTACTTGAGGTAAGCAATA 22929
133 ..... 133
22930 TCTTGATCTTGAGATGACATTTTAACTGATTTTGTGTATATTT 22979
134 .....As 134
22980 TTAATAACAAATGCTTAACTTTTACCTTTTACACCTTATTTTAGAA 23029
134 ntleAlatrpVallySerTyrAsnIleCysPheGluLeuGluAspSera 151
|||||
23030 CATTCCTGGGTCAGTCATATAACATATGCTTTGAGTTAGAGATAGCA 23079
151 snGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIle.Asn.. 166
|||||
23080 ATGAATTTTCCACCAGCTATACAGAACCTTTTTCAGTTATATAAGTAA 23129
166 ..... 166
23130 GTTATTTTATTAAGTATGATACATTTAAAAAAGGTAACAAATTTTGAAC 23179
166 ..... 166
23180 TTGTAGTGAATCTGCATAAACTGTTACAAATGATATTTAGAGGCTGAGA 23229

```

```

166 ..... 166
23230 GATGTAACCTTTTAAACACACTTTATTTTATACATGTTTCAAAAATGAATA 23279
166 ..... 166
23280 GAGCTTAATTTTGTATATATATATTAATTTATTTGTTAGAGAGTGGCATAAG 23329
166 ..... 166
23330 GCCAGTGTAATAATTTTGGGATTGCATTTATTTGCTATATAAACATACAT 23379
166 ..... 166
23380 ACTGACTTCTTAATGATATGCTATTTGAGCTTTTGTGTATGCTTTAA 23429
166 ..... 166
23430 AATTATTTACTGGAATTAACCTTCAGAGCAATTCATTTTCTGAATTTAAG 23479
166 ..... 166
23480 GCAGAACTGAAGTTGCTTTTGTATGTAATTTACTTTTGTGAGACTGGT 23529
166 ..... 166
23530 TAAAAATTAGGGAAAAAATCTAAGAAATGTAGTGAATAAATAATAATTTGTA 23579
166 ..... 166
23580 TACTTTTGTATGATGTTGGAGAGAGGGAAGCTTTAAATGAATATACATAG 23629
166 ..... 166
23630 AATTTTATTTAATATTTTACTATTTTGTAAAGCATAAATGCTTGTCA 23679
166 ..... 166
23680 GTGTCATCTTAAGAGTAATAATCCTGTATATCTCAGTTGTAAATGAAACAAA 23729
167 .....AsnGlyHisAsnGlnLysValHisMetHisM 177
|||||
23730 ATCTGTTTGAATTCAGCAATGGCCACATGAGAAAGTCCATATGCACA 23779
177 evValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGln 193
|||||
23780 TGTAGACCTTATGAGCTCTATTTTGTGAAGGTGATACAGTGTCTCAG 23829
194 GluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLys 208
|||||
23830 GAGCTTTGGATACGGTTTGTAGTAATCTGGTACCTGCTCATAAG 23874

```

```

seq_name: gb_htg20:AL353724

```

seq_documentation_block:

```

LOCUS AL353724 163642 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-448113, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL353724
VERSION AL353724.3 GI:9930876
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163642)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Aug 27, 2000 this sequence version replaced gi:9926534.

```



```
/note= LI element fragment
```



```

misc_feature 1..2333
/note="assembly_fragment:02826
fragment_chain:1"
2434..4472
/note="assembly_fragment:00159
fragment_chain:1"
4573..6623
/note="assembly_fragment:00292"
6724..8757
/note="assembly_fragment:02097"
8858..11498
/note="assembly_fragment:02560"
11599..13673
/note="assembly_fragment:03440"
13774..16148
/note="assembly_fragment:03977"
16249..195380
/note="assembly_fragment:04183"
55503 a 40458 c 40664 g 58048 t 707 others
ORIGIN

alignment_scores:
  Quality: 401.50      Length: 361
  Ratio: 3.788         Gaps: 2
Percent Similarity: 29.363 Percent Identity: 29.086

alignment_block:
US-09-512-581-2 x AL358892 ..
Align seg 1/1 to: AL358892 from: 1 to: 195380

103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAs 119
|||||
170504 TTAAGGATATATTTATGTTTATAACAAGGCAACTAAAGGGACTAGAAGA 170553

119 pThrLysSerProGlnPheAsnArgTyPheTyrlLeuLeuGlu..... 133
|||||
170554 TACAAGAGCCCTCAATTTAATAGGTATTTTATTACTTGAGGTAAGCA 170603

133 ..... 133

170604 ATACATTATTACTTGAGTTGATTTCTTTTCTTCTATATATCTTTAGC 170653

134 .....AsnIleAlaTr 137
|||||
170654 AACCAATATATTACTTTCCTCCCTACTTCTTTCTTTTGAACATTCGATG 170703

137 pValLysSerTyAsnIleCysPheGluLeuGluAspSerAsnGluIleP 154
|||||
170704 GGTAAATCATATACATATGCTTTTGAGTTAGAAGATGAATGAATCT 170753

154 heThrGlnLeuTyArgThrLeuPheSerValIle..... 165
|||||
170754 TTACTCAACTTTACAGAACATTTATCTCAGTTATAAGTTAAGTTGATTT 170803

165 ..... 165

170804 TAAATAACATTTCTTTTAAATGTAAAGCTTTTCTCTTTACTATGAATCC 170853

165 ..... 165

170854 ACATAAAGCTGTATAAATATTGCTTAGGATTTATAAATTAACCTTACAAT 170903

165 ..... 165

170904 TATTCTCTATATAGTTCTAAATGAATAGATTTTATTATATATATTGTT 170953

165 ..... 165

170954 AGAAGAGCAATAGGCTAAGTTTGGGATTTATTTCTTTCTGATTACAT 171003

165 ..... 165

```

```

171004 ACTATGTTAACTTTTGTGTTCATTGAGATGTTTCCCATGAGCTCTGG 171053
165 ..... 165
171054 CTGTGCTAGAACTCTCTGCAACACAGGCTGGCCTCAGACTCAGAGCCT 171103
165 ..... 165
171104 GCTTCTGCTTCCCATGTTGCTGAGATCAATGGCATGGCATTGACCATCATCTGA 171153
165 ..... 165
171154 CTTATTGTGACTTTTAAATATATATCATGTTGTTGGAAATTTTGTTTTATC 171203
165 ..... 165
171204 TATAAATGTAATAGAAATTTTAAATTTAAGAACTAAGATTGCATTTC 171253
165 ..... 165
171254 TTTAATTACTTTGGTGAAGAATGCTAAATATTAGGGGAAAGTAATTTA 171303
165 ..... 165
171304 GTAATATAGTTTAAATAATACATTTTATTATTGTAAGACAGAGAGAAA 171353
165 ..... 165
171354 GAAGCTTGACAGTAGTACAGATTTATCTCACCACITTCATATAAGTATAT 171403
165 ..... 165
171404 GTTACTACTACCATATCTATCTCAGTTGAAATAAAATAAAGTGTGGG 171453
166 .....AsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuM 181
|||||
171454 TTACAGCAATGGCCCAATCAAGAAAGTTTCATATGCACATGGTGGACCTCA 171503
181 etSerSerIleCysGluGlyAspThrValSerGlnGluLeuLeuAsp 197
|||||
171504 TGAGCTCTATCATTTGTGAAGGTATACCGTATCTCAGGAGCTCTTAGAT 171553
198 ThrValLeuValAsnLeuValProAlaHisLys 208
|||||
171554 ACAGTTTTAGTAAATCTGTTACCTGCCCAATAG 171586
seq_name: gb_pr8:HS267P19
seq_documentation_block:
LOCUS HS267P19 113704 bp DNA 22-NOV-1999
DEFINITION Human DNA sequence from cosmid 267P19, BRCA2 gene region chromosome
13q12-13 contains polymorphic CA repeat.
ACCESSION 275889
VERSION 275889.1 GI:1430780
KEYWORDS 13q12-13; repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113704)
Odell,C.
Direct Submission
Submitted (04-JUL-1996) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
IMPORTANT:
This sequence is not the entire insert of clone 267P19. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,

```



```
904 yrglnLeuCysAlaLeuAlaIleAsn..... 912
:|||||
92230 TTCAGCTCTGTCACCTTGTATTAAATAGTAACAATCTTATTTTGTG 92181
912 ..... 912
92180 AATTACTTATCTCAAGATTGAATACATTACTTTTCTATGATACAAATC 92131
912 ..... 912
92130 CATTATAATGTGAATATATGACAAAAGTCACTACTATTATTACCTGTT 92081
912 ..... 912
92080 TAACTGCTTCCAAATTAATTAATTAATTAATTAATTAATTAATTAAT 92031
912 ..... 912
92030 TTTTAAATGAAGACACATAAAATTAATTAATTAATTAATTAATTAAT 91981
913 ..... Asp.GluCysTyr 916
91980 AATAGGTGTCGAATTAAGTTTCTGCTGCTGCTAGGATGAGTGTAC 91931
917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLe 933
91930 CAAGTAAGGCAGATATTTGCTCAGAGCTGCATAAGCCACTTGTGAAGTT 91881
933 uArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 950
91880 ACTGCTCCCAATGGAGTATATGGCATCTTGGCTTGTGTGCCAAAGATC 91831
950 roValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
91830 CTGTGAAGGAGAGAGACACACGACGACAAATGTTTACTGAAAATATC 91781
967 AsnValArgArgGluTyrLeuLysGlnHisAlaAla.Val..... 979
:|||||
91780 AGTATAGCAGGAATACATTAAGCAGAACTCTATGGCTACTGGTAAGTA 91731
979 ..... 979
91730 ACTCAGAGTTCTCCGTGTCAAATATATTTTGTAAATGTATCCATTTTA 91681
979 ..... 979
91680 GGTTTTTTTTTCTGTGTCATGCTTTCTATTATAGATTATAATAAACTG 91631
979 ..... 979
91630 GGAGTTTATAGAATGCTGTCATTTAGTTATGATGTACTTAAAACTGA 91581
979 ..... 979
91580 ACTCCTGAGTCCCTATTATGCTGCATTAACTCGCAAACTGCATTTTTC 91531
979 ..... 979
91530 CTATGGGATCAGGAGTGTAGTTAAATTTGAACATATTGCTCGGATGTCAG 91481
979 ..... 979
91480 ACTAAATAATTTTCTTCTAGCTTTGCCAGGACCTGTGTTTAAATGGTT 91431
979 ..... 979
91430 TAATATCTGATACCTTTGACTTCAGTATCTTTCAAAAGTGAATACAACT 91381
979 ..... 979
91380 GCATACCAGGTACCAGAAAAGAAATTTGAAAAGTGAAGTCTAAAGTATTT 91331
979 ..... 979
```

```
91330 TAGTATCCAGGTATCATGTTGTGCTTAATTTAGAAATAACTGAGGAAAA 91281
979 ..... 979
91280 AATGAAAACTTAAGTGTGTAGAACTCAGTCCACTAAGGCATTTTATAAAG 91231
980 ..... SerGluLysLeuLeuSerLeuLeuPr 988
91230 GTTGTGTTTAAATAATGTTTACTTTCAGAGAAATATTATCAGCTGTGCC 91181
988 oGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1005
91180 TGAATATGTAGTTCATACATGATTCACCTGTAGCCCATGATCCAGATT 91131
1005 yrvValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeu 1021
91130 TTACAAGATCACAAAGATGTTGATCAGCTGCTGATATCAAGAGTAAGTC 91081
1022 TrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsn 1035
91080 ..... TGTGTGTTTTCAGAACTCTAAAAAT 91054
```

seq_name: gb_htg21:AL358892

seq_documentation_block:

LOCUS AL358892 195380 bp DNA HTG 06-MAR-2001

DEFINITION Mus musculus chromosome 5 clone RP21-417G6, *** SEQUENCING IN PROGRESS ***, in unordered pieces.

ACCESSION AL358892

VERSION AL358892.12 GI:11493289

KEYWORDS HTG; HTGS-PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Wall, M

Direct Submission

Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequests@sanger.ac.uk

On Nov 30, 2000 this sequence version replaced gi:11414592.

COMMENT

----- Genome Center

Center: UK Medical Research Council

Center code: UK-MRC

Web site: http://mrcseq.har.mrc.ac.uk

Contact: mouseq@har.mrc.ac.uk

----- Project Information

Center project name: dm417G6

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 190612 bases at least Q40

Consensus quality: 192224 bases at least Q30

Consensus quality: 193166 bases at least Q20

Insert size: 194680; sum-of-contigs

Insert size: 191947; 3.7% error; agarose-fp

Quality coverage: 7.33x in Q20 bases; sum-of-contigs Quality

coverage: 8.85x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

Source

Location/Qualifiers

1. 195380

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="5"

/clone="RP21-417G6"

/clone_lib="RPCI-21"


```

* 1519 2614: contig of 1096 bp in length
* 2615 2714: gap of 100 bp
* 2715 4205: contig of 1491 bp in length
* 4206 4305: gap of 100 bp
* 4306 7004: contig of 2699 bp in length
* 7005 7104: gap of 100 bp
* 7105 9239: contig of 2135 bp in length
* 9240 9339: gap of 100 bp
* 9340 11600: contig of 2261 bp in length
* 11601 11700: gap of 100 bp
* 11701 13258: contig of 1558 bp in length
* 13259 13358: gap of 100 bp
* 13359 16016: contig of 2658 bp in length
* 16017 16116: gap of 100 bp
* 16117 18620: contig of 2504 bp in length
* 18621 18720: gap of 100 bp
* 18721 21036: contig of 2316 bp in length
* 21037 21136: gap of 100 bp
* 21137 23167: contig of 2031 bp in length
* 23168 23267: gap of 100 bp
* 23268 26524: contig of 3257 bp in length
* 26525 26624: gap of 100 bp
* 26625 28953: contig of 2329 bp in length
* 28954 29053: gap of 100 bp
* 29054 32709: contig of 3656 bp in length
* 32710 32809: gap of 100 bp
* 32810 35161: contig of 2352 bp in length
* 35162 35261: gap of 100 bp
* 35262 36494: contig of 1233 bp in length
* 36495 36594: gap of 100 bp
* 36595 40718: contig of 4124 bp in length
* 40719 40818: gap of 100 bp
* 40819 44641: contig of 3823 bp in length
* 44642 44741: gap of 100 bp
* 44742 48769: contig of 4028 bp in length
* 48770 48869: gap of 100 bp
* 48870 53265: contig of 4396 bp in length
* 53266 53365: gap of 100 bp
* 53366 57227: contig of 3862 bp in length
* 57228 57327: gap of 100 bp
* 57328 60723: contig of 3396 bp in length
* 60724 60823: gap of 100 bp
* 60824 64959: contig of 4136 bp in length
* 64960 65059: gap of 100 bp
* 65060 69229: contig of 4170 bp in length
* 69230 69329: gap of 100 bp
* 69330 75479: contig of 6150 bp in length
* 75480 75579: gap of 100 bp
* 75580 84431: contig of 8852 bp in length
* 84432 84531: gap of 100 bp
* 84532 92387: contig of 7856 bp in length
* 92388 92487: gap of 100 bp
* 92488 99945: contig of 7458 bp in length
* 99946 100045: gap of 100 bp
* 100046 112666: contig of 12621 bp in length
* 112667 112766: gap of 100 bp
* 112767 126095: contig of 13329 bp in length
* 126096 126195: gap of 100 bp
* 126196 145450: contig of 19255 bp in length.

```

FEATURES source

```

1..145450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-168F19"
/clone_lib="RPC1-11 Human Male BAC"
1..1418
/note="assembly_fragment"
1519..2614
/note="assembly_fragment"
2715..4205
/note="assembly_fragment"

```

```

misc_feature
misc_feature
misc_feature

```

```

misc_feature
4306..7004
/note="assembly_fragment"
misc_feature
7105..9239
/note="assembly_fragment"
misc_feature
9340..11600
/note="assembly_fragment"
misc_feature
11701..13258
/note="assembly_fragment"
misc_feature
13359..16016
/note="assembly_fragment"
misc_feature
16117..18620
/note="assembly_fragment"
misc_feature
18721..21036
/note="assembly_fragment"
misc_feature
21137..23167
/note="assembly_fragment"
misc_feature
23268..26524
/note="assembly_fragment"
misc_feature
26625..28953
/note="assembly_fragment"
misc_feature
29054..32709
/note="assembly_fragment"
misc_feature
32810..35161
/note="assembly_fragment"
misc_feature
35262..36494
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature
36595..40718
/note="assembly_fragment"
misc_feature
40819..44641
/note="assembly_fragment"
misc_feature
44742..48769
/note="assembly_fragment"
misc_feature
48870..53265
/note="assembly_fragment"
misc_feature
53366..57227
/note="assembly_fragment"
misc_feature
57328..60723
/note="assembly_fragment"
misc_feature
60824..64959
/note="assembly_fragment"
misc_feature
65060..69229
/note="assembly_fragment"
misc_feature
69330..75479
/note="assembly_fragment"
misc_feature
75580..84431
/note="assembly_fragment"
misc_feature
84532..92387
/note="assembly_fragment
clone_end:T7
vector_side:left"
misc_feature
92488..99945
/note="assembly_fragment"

```

```

alignment_scores:
    Quality: 466.00    Length: 432
    Ratio: 3.149      Gaps: 5
    Percent Similarity: 34.259    Percent Identity: 28.704

```

```

alignment_block:
US-09-512-581-2 x AC068352/rev ..

```

```

Align seg 1/1 to reverse of: AC068352 from: 1 to: 145450

871 SerLysProAspMetSerArgLeuAlaAlaGlySerAlaIleVa 887
||||| ||||||| ||||||| ||||||| ||||||| |||||||
92330 AGTAATCTGATATGCTCGCTGCGATTAGCTGCTGGTAGTGCATAAT 92281

887 llyLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnT 904
||||| ||||||| ||||||| ||||||| ||||||| |||||||
92280 GAAGCTTGCTCAGGAACCTTGTTACCATGAATATTATCCCCAGAACACT 92231

```


4355 TATATCTGAAGTGTGTTACT.....CTTTGCTGCCATTGA 4389
 910AlaIleAsnAspGluCysTyrGlnValar 919
 4390 ATAATATCTGGAATAATGTGCTTTACAGAGTAATGCTATCAAGTAAG 4439
 919 gGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuP 936
 4440 ACAAGTGTTCGCCAGAACTTCACAAAGCCCTTCCTCGTTTACGGCTTC 4489
 936 roLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysaspProValLys 952
 4490 CACTTGAGTATATGCAATCTGTCCCTTTGTGCAAAAGATCTGTAAAG 4539
 953 GluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValar 969
 4540 GACAGAGAGCTCATCTAGGCAATGTTGGTGAAATAATAATGTAAG 4589
 969 gArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuSerL 986
 4590 GCGGGAGTATCTGAAGCAGCATCGCTGTTAGTGGTAAGCATATAAGAA 4639
 986 euLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAsp 1002
 4640 AATGGAAAGGATACCTTTTTCAGCGCTGCTAGTTTCAGTTTATAGAATA 4689
 1003 ProAspTyrValLysValGlnAspIleGluGlnLysAspValLysG 1019
 4690 GCATGATATATG.....ATATTACTG..... 4701
 1019 ucLysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnS 1036
 4702ATATTACTG..... 4710
 1036 erHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
 4710 4710
 1053 AlaGlnProAspAspAlaLysMetAsnGluLysLeuTyrThrValCy 1069
 4711TTATTATTATTG 4724
 1069 sAspValAlaMetAsnIleIleMetSerLysSerThrTyrSerLeuG 1086
 4725 TGATTACAAATT..... 4737
 1086 luSerProLysAspProValLeuProAlaArgPhePheThrGlnProAsp 1102
 4738TTATTACCAANNNNNNNNNNNNNNNNNNNNNNNN 4770
 1103 LysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLysSerPh 1119
 4771 NNN 4820
 1119 ePheThrProGlyLysProLysThrThrAsn.....ValL 1131
 4821 NNN 4870
 1131 euGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr 1147
 4871 TCGGCTCTNANAAAAACA.CTTTTCATCAGCAGCAAGCAATCTCAGACC 4919
 1148 LysSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPr 1164
 4920 AAATCATCAGTAATGGAACCTGTGAAGCAATGCAAGCAGCAGCTCAATCC 4969
 1164 oSerSerProGlyArgIleLysGlyArg 1173
 4970 AAGCTCTCTGGGAATAAAGGGAGG 4997

seq_name: gb_htg14:AC068352

seq_documentation_block:

LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

AC068352 145450 bp DNA HTG 25-JUN-2000
 Homo sapiens chromosome 4 clone RP11-168F19 map 4, WORKING DRAFT
 SEQUENCE, 31 unordered pieces.
 AC068352
 AC068352.2 GI:8705054
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McSurk,A., McKernan,K., McPheters,R.,
 Meidirm,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 COMMENT

Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 25, 2000 this sequence version replaced gi:7677723.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10143

Center clone name: 168_F19

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 126425 bases at least Q40

Consensus quality: 135577 bases at least Q30

Insert size: 172000; agarose-fp

Quality coverage: 3.4 in Q20 bases; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1418: contig of 1418 bp in length
 * 1419 1518: gap of 100 bp


```

524 .....AlaIlePheSerLy 528
774 TCTGTTGCTTGGGATAAAGTTGAACACCTTATAGCCCTTTATAGTTT 2823
528 sValMetVal...IleThrArgAsnLeuProAspPro..... 539
2824 GCTCTCAGCCTGAGTTTCAAGCTTTCTTCTGCTCCCTATTATGAATTT 2873
540 .....GlyLysAlaGlnAspPheMetLys.....LysPheThr 550
2874 GTGTTTCATTACTTCGATTACTTGAATTTGTGTACAGTCTTTTTC 2923
551 GlnValLeuGluAspGluLysIleArgLysGlnLeuGluValLeuVa 567
2924 TCTGTTGTCACTATCTATCGAAATTCCTCT.....CT 2958
567 lSerProThrCysSerCysLysGlnAlaGluGly..... 578
2959 TTCCAGACAACAACATGAACAACCAAGGAGACATGGTTGACACCTT 3008
579 .....CysValArg.....GluIleThrLysLysLeu 587
3009 TTGAATGCTTTAAATACAGTTTGGGGTTACCACATCACTGTGACCCTC 3058
588 .....GlyAsnProLy 591
3059 TTTCTTTTTTTTTCCTTCTTACGCTGGAAATGTGAGGGTCTGCAGTC 3108
591 sGlnProThrAsnPro...PheLeuGluMetIleLysPheLeuGluVa 607
3109 ACATCCCACTTCTGCGTTTACTGTGCGTATTGTCGAGATTACTTAATC 3158
607 rgIleAlaProValHisIleAspThrGluSerIleAlaLeulle... 622
3159 ATTTGTCCTT.....AGTTTCCTTATTTCG 3184
622 ..... 622
3185 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3234
622 ..... 622
3235 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3284
623 .....LysGlnValAsnLysSerIleAspGlyThrAlaAsp... 634
3285 ACATAACTGGTGCAGAAACCGCAACTCGTTAGAGAAAACCTGCTGATTT 3334
635 AspGluAspGluGly...ValProThrAspGlnAlaIleArgAlaGlyLe 650
3335 GATTGAGATCATGGGAATTTGCTAGG..... 3361
650 uGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSera 667
3362 .....AAAGTGAAGTTT...CACCCCTTAGATGAACAACA 3398
667 lAcLutThrPheGluSerLeuLeu...AlaCysLeuLysMetAspGluL 683
3399 CTGCTCCATTGACCTCTGCTCATGAAGTGATG..... 3432
683 ysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIle 699
3433 .....TGTGGAAGGGAGATT 3447
700 GluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisH 716
3448 GAGAAAAG.....CACTC 3461
716 sLysSerLysLysGly.....ProProArgGlnAlaL 727
3462 TAAACTGAAAAAGAAAAAGAACTACTCTGTTTTCCTCGAAGCAATTTA 3511
727 ys.....TyrAlaIleHis.CysIl 733
3512 AAATGACCAGGCGCAGTCTGTTTTTAGATGAATTCCTTTTACATGCT 3561
733 ehIsAlaIlePheSerSerLysLutThrGlnPheAlaGlnIlePheGluP 750
3562 ACATAGCATTTCATGAAAAAAAACCCCTAGATCATCTTAAATG 3611
750 ro..... 750
3612 ATTGATAAAAGATAAAATATGATACATCAATGATGAATACCATCATTA 3661
751 .....LeuHisLysSer...LeuAs 756
3662 CATTATAATACTAGTAGTACATATAAATCTTCATAAATATTAAGTAAA 3711
756 pProSerAsnLeuGluHisLysIleThrProLeuValThrIleGly... 771
3712 AACATCAAGATACCTAACTTAATTAATTAATTAATTAATTAATTAATTA 3761
772 .....HisIle 773
3762 CCACAGCATATACAAAAAGTAGGAACATCTTAGAGTGGCAACATAAA 3811
774 AlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerTrpVala 790
3812 GTGCAATTTCTTCGATTTCATGCTTTCAATAATGCTGAGTTATTGTT 3861
790 lThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
3862 ACACCTTATTATTAAG...CTACCAAAAACTTTGAGTACACAGACCCA 3908
807 LysThrThrLysLeuTrpVal...ProAspGluGluValSer..... 819
3909 CACTCTCTGTAATGTGGTATTACCTTTTGCAGATATAAACATTTATCA 3958
819 ..... 819
3959 CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4008
819 ..... 819
4009 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4058
820 .....ProGluThrMetValLysIleGlnAlaIl 829
4059 NAATACTATTNGCAAAACCCCCACACTTGTCTTANAGAAATCCCC..... 4103
829 eLysMetMetValArgTrpLeuLeuGlyMetLys..... 840
4104 .....TATATGAGATGGTGTAAAGCCTTAATTTTGTATCTATTATTA 4146
841 .....AsnAsnHis 843
4147 AGTTAGGCGAGATTATATTTACTAGTTTTTTTCTGTTTTTAAAGTATGT 4196
844 SerLysSerGlyThrSerThrLeuArgLeuLeu..... 854
4197 TTGTTTTCTAACACATCTAACTGAGTACTATTAGGCGCCAGATGAATTT 4246
855 .....ThrThrIleLeuHisSera 861
4247 GCAGCTTCAAAATGAACCTTTAGTCACACATCTCCTCTTATGAGCAGTG 4296
861 spGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
4297 AAGCAATTTATCTAGAT..... 4313
878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCy 894
4314 ...AAACTGCTACTGTTTACTCTTACTTAAATAGCAAT.....TT 4354
894 styrHisGluIleIleThrLeuGluGlnThrGlnLeuCysAlaLeu... 909

```



```

* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length
* 39913 40012: gap of 100 bp
* 40013 40712: contig of 700 bp in length
* 40713 40812: gap of 100 bp
* 40813 41514: contig of 702 bp in length
* 41515 41614: gap of 100 bp
* 41615 42294: contig of 680 bp in length
* 42295 42394: gap of 100 bp
* 42395 43094: contig of 700 bp in length
* 43095 43194: gap of 100 bp
* 43195 43915: contig of 721 bp in length
* 43916 44015: gap of 100 bp
* 44016 44740: contig of 725 bp in length
* 44741 44840: gap of 100 bp
* 44841 45533: contig of 693 bp in length
* 45534 45633: gap of 100 bp
* 45634 46330: contig of 697 bp in length
* 46331 46430: gap of 100 bp
* 46431 47137: contig of 707 bp in length
* 47138 47237: gap of 100 bp
* 47238 47927: contig of 690 bp in length
* 47928 48027: gap of 100 bp
* 48028 48709: contig of 682 bp in length
* 48710 48809: gap of 100 bp
* 48810 49509: contig of 700 bp in length
* 49510 49609: gap of 100 bp
* 49610 50331: contig of 722 bp in length
* 50332 50431: gap of 100 bp
* 50432 51135: contig of 704 bp in length
* 51136 51235: gap of 100 bp
* 51236 51981: contig of 746 bp in length
* 51982 52081: gap of 100 bp
* 52082 52783: contig of 702 bp in length
* 52784 52883: gap of 100 bp
* 52884 53580: contig of 697 bp in length
* 53581 53680: gap of 100 bp
* 53681 54398: contig of 718 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..54398
      /organism="Homo sapiens"
alignment_scores:
  Quality: 554.50      Length: 1213
  Ratio: 1.160        Gaps: 53
  Percent Similarity: 39.406      Percent Identity: 25.144
alignment_block:
  US-09-512-581-2 x AC068224
Align seg 1/1 to: AC068224 from: 1 to: 54398
283 SerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMe 299
|||||
1956 AGCAATGATGATGAGGAGCGCTACAAAGTTGTTAAACTACTGGCAAAAT 2005

```

```

299 tPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpG 316
|||||
2006 GTTGGGGCAAGGATTCAGAAATGGCTTCTCAAAACAAGCCACTTTGGC 2055
|||||
316 lncystyrLeuGlyArgPheAsnAspIle..... 325
|||||
2056 AGTGCTACTTTGGCAGGTATATGATTTGGTTCCCATTTAAAGTAATA 2105
|||||
326 .....HisValProlleArgLeuG 332
|||||
2106 CATATTCTACTGACCGTTTTTTTAAAGAAACATTATCCA...CATTTAAA 2152
|||||
332 uCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaL 349
|||||
2153 CTGTGTT.....CfCTGGAATCATAAAATGAAGGAT 2184
|||||
349 ysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAla 365
|||||
2185 TTTTATGTTGAAAGTTTTCGAAAGTT..... 2210
|||||
366 lIeArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIl 382
|||||
2210 .....HisLeuLeuAsnPheVal.Arg 394
|||||
382 eLeuLeuValAsnAsp..... 2211
|||||
2211 .....GTTAATGACGTGGCTATATTATATTACTCTTTTACTCTTTA 2253
|||||
395 GluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGl 411
|||||
2254 TCCTCAGTAATTTTAAAGATAATACATTTTAAAGCTATCTCTTATGGT 2303
|||||
411 yLeu.....AlaGlnIleTyrLysTyrAlaLeuGlnSerA 424
|||||
2304 ATTATGACCATGAAGTTAAATATTATTACTTTAAGGCAGTGATTAAATCCA 2353
|||||
424 laAlaGlyLysAspAlaAla..... 430
|||||
2354 ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2403
|||||
430 ..... 430
|||||
2404 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2453
|||||
431 .....LysGlnIleAlaTrpIleLys.....AspLysLeuLeuH 442
|||||
2454 NGTCTGNNNGNGTAAACATGGTTAAAGCTGTACCTCGTTANAGTTAAC 2503
|||||
442 isIleTyrTyrGlnAsnSerIleAspArgLeuLeuValGluArgIle 458
|||||
2504 ACTNCTAC.....TTAGAGGCATTA 2523
|||||
459 PheAlaGlnTyrMetValProHisAsn.....LeuGluTh 470
|||||
2524 TTCATTCTTTTTCCTCTCTCAACCGTCACCAAAATTCGTGTTAACT 2573
|||||
470 rThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuA 487
|||||
2574 TTCTCCAATGCTCTTGAACATTTATCTTCTACTCTCTCTCTCTATAT 2623
|||||
487 snAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn..... 500
|||||
2624 TTTCACTTGGATGCTGTATACATTTGCTCCCTGCTGCTGTTGTAATGTTCTGT 2673
|||||
501 .....Leu.LeuArgHisGlnValLysAspLeuLeuAspLeuI 513
|||||
2674 GCCTCTGACCCACTGCTGCTGTATACATTCATCAAAATTTATCTCTC 2723
|||||
513 lelysGlnProLysThrAspAlaSerValLys..... 523
|||||
2724 TAAACACAGATTAGATCATCATCAGTCCCTCTTCTGAAATTTCTGTTGATTC 2773
|||||

```


4621 ACCCTTCCA 4629

seq_name: gb_hgt14:AC068224

seq_documentation_block:

LOCUS AC068224 54398 bp DNA HTG 30-APR-2000
 DEFINITION Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
 SEQUENCE SAMPLING.

ACCESSION AC068224
 VERSION AC068224.1 GI:7671284
 KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 54398)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 3, clone RP11-660H19

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 54398)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boquslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
 Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Maylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10161

Center clone name: 660_H_19

* NOTE: This record contains 68 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 685: contig of 685 bp in length
 * 686 785: gap of 100 bp
 * 786 1513: contig of 728 bp in length
 * 1514 1613: gap of 100 bp
 * 1614 2354: contig of 741 bp in length
 * 2355 2454: gap of 100 bp

2455 3184: contig of 730 bp in length
 * 3185 3284: gap of 100 bp
 * 3285 3959: contig of 675 bp in length
 * 3960 4059: gap of 100 bp
 * 4060 4746: contig of 687 bp in length
 * 4747 4846: gap of 100 bp
 * 4847 5564: contig of 718 bp in length
 * 5565 5664: gap of 100 bp
 * 5665 6360: contig of 696 bp in length
 * 6361 6460: gap of 100 bp
 * 6461 7132: contig of 672 bp in length
 * 7133 7232: gap of 100 bp
 * 7233 7937: contig of 705 bp in length
 * 7938 8037: gap of 100 bp
 * 8038 8749: contig of 712 bp in length
 * 8750 8849: gap of 100 bp
 * 8850 9522: contig of 673 bp in length
 * 9523 9622: gap of 100 bp
 * 9623 10356: contig of 734 bp in length
 * 10357 10456: gap of 100 bp
 * 10457 11181: contig of 725 bp in length
 * 11182 11281: gap of 100 bp
 * 11282 11975: contig of 694 bp in length
 * 11976 12075: gap of 100 bp
 * 12076 12785: contig of 710 bp in length
 * 12786 12885: gap of 100 bp
 * 12886 13557: contig of 672 bp in length
 * 13558 13657: gap of 100 bp
 * 13658 14347: contig of 690 bp in length
 * 14348 14447: gap of 100 bp
 * 14448 15132: contig of 685 bp in length
 * 15133 15232: gap of 100 bp
 * 15233 15912: contig of 680 bp in length
 * 15913 16012: gap of 100 bp
 * 16013 16731: contig of 719 bp in length
 * 16732 16831: gap of 100 bp
 * 16832 17506: contig of 675 bp in length
 * 17507 17606: gap of 100 bp
 * 17607 18339: contig of 733 bp in length
 * 18340 18439: gap of 100 bp
 * 18440 19109: contig of 670 bp in length
 * 19110 19209: gap of 100 bp
 * 19210 19900: contig of 691 bp in length
 * 19901 20000: gap of 100 bp
 * 20001 20693: contig of 693 bp in length
 * 20694 20793: gap of 100 bp
 * 20794 21464: contig of 671 bp in length
 * 21465 21564: gap of 100 bp
 * 21565 22245: contig of 681 bp in length
 * 22246 22345: gap of 100 bp
 * 22346 23049: contig of 704 bp in length
 * 23050 23149: gap of 100 bp
 * 23150 23857: contig of 708 bp in length
 * 23858 23957: gap of 100 bp
 * 23958 24685: contig of 728 bp in length
 * 24686 24785: gap of 100 bp
 * 24786 25499: contig of 714 bp in length
 * 25500 25599: gap of 100 bp
 * 25600 26302: contig of 703 bp in length
 * 26303 26402: gap of 100 bp
 * 26403 27108: contig of 706 bp in length
 * 27109 27208: gap of 100 bp
 * 27209 27910: contig of 702 bp in length
 * 27911 28010: gap of 100 bp
 * 28011 28720: contig of 710 bp in length
 * 28721 28820: gap of 100 bp
 * 28821 29484: contig of 664 bp in length
 * 29485 29584: gap of 100 bp
 * 29585 30328: contig of 744 bp in length
 * 30329 30428: gap of 100 bp
 * 30429 31111: contig of 683 bp in length
 * 31112 31211: gap of 100 bp
 * 31212 31931: contig of 720 bp in length


```

VERSION      L03200.1  GI:168024
KEYWORDS     DNA-damage sensitive; mitotic block.
SOURCE       Emericella nidulans cDNA to mRNA.
ORGANISM     Aspergillus nidulans
             Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
             Eurotiales; Trichocomaceae; Emericella.
REFERENCE    1 (bases 1 to 4885)
AUTHORS      Denison,S.H., Kafer,E. and May,G.S.
TITLE        Mutation in the bimd gene of Aspergillus nidulans confers a
             conditional mitotic block and sensitivity to DNA damaging agents
JOURNAL      Genetics 134 (4), 1085-1096 (1993)
MEDLINE      93387663
FEATURES     Location/Qualifiers
             1..4885
             /organism="Aspergillus nidulans"
             /db_xref="taxon:5072"
             133..4653
             /gene="bimd"
             133..4653
             /gene="bimd"
             /codon_start=1
             /protein_id="AA03063.1"
             /db_xref="GI:168025"
             /translation="MPSRTROGQPSVATEDVEPEEPSSGGLRQLKFNPLELSWRVGRSAIP
             IADLRQLTQAEQPNVQEETIEKESLRKVSQELATQAOLLAHKDKGVRAWATCCIVD
             VLRLCAPDAPFTANOLKIDFTICVSIIPALGDPSPNPYNAOHYVILNSLAEVKSIYLM
             TLDHPDTLIVPLFTSCFDIVAGSAKASTQEPVAKNVEDMTFLVTVIDESPVLAPD
             VYDVIAQFLRVDPVLDPGPKGKKPEQVDEKQETLLLLKDYPAAYNMAKALCQACP
             EMTSHISQIFNNVIIDASATQNGSPKQARTNLDDSDGEGDIKLSKAHLIRE
             LWRACPDVLQNYIPOLEAELSAESYLRLLATQTIGDLTSFGVAGPPPLPMPFVAY
             POKLDDYARSLPQNPVLPFPKPPFQSAHSSAYDSFLSRRLDKSASVRASWRPTAIG
             RIILTSAGSGSLSDNEOETLITHLSMLRDARVRLAAVEAVGTFLSHIVNKLGV
             GSVQDSLLFTILAEVRKDRKSQVREHATPKVLARAWAVASGDIERSHEOVTPLLKEAP
             SRILDATYNDIEIHSIDRAMFEILLPLSYPIPKPLSRSSSSORLKSQAAEPE
             SEADVIRIKVRRLILVGLDEKAKVFAMQKRVSLRTAVTVYLQACEEYNGWYE
             KNKQIKQALTRIVDALAKTFDPARTSADLWKFAKHDRRGYQLIRFAMAAYSRYT
             VTKAKELARLQSSNNTILHETLTLLYRCSIVFNRSHPAIMSISRSDENGLAAP
             AHMKLEISNPVELEAOVQKDELAQAPKATVSAAGTEELIKACSGFAKKLPS
             KLPKERRKFOALVDVALHSPSPRAAKHVALILMAYTDKKNMVAKDLVEKCSKCYVDS
             ERFPLKATLSQLNLLAPREADESDAIKISYNOILLNTRSPFNSGFYFSDQVDDE
             TKAEWALKILVNRUKRQSDSDDFRAHAEPVYDITLANKLVNSELGSKKIDTPTAQ
             KSRLLRANLSLLKSCSHALCEQLTPODFNSIALVAQDPLPEVRSGFINQLKKLV
             OTRLGARWYVYVLIAPFQVGLKDLSTLWLSRAAFSQOTNGKKGKEQVMEALF
             SLLLSLAYHPDYPPADLDESKLDLDTFYARYILFYSAVANEHNLIFLFIHAORVK
             QARDGTTKSDMSRRLHTLSDLAQATIRFADVYSOORFEGGAGGVNLLQYPPGVK
             VPSSFPADNASHREAEVADTTFLEDAEDRLDRIVRATMRFTKNSQAGAKKRTYD
             TQPSFDAANAARKANSESSRRKSGSAVGFKMPKRRKTKTKGDDWSSDGEAAGNV
             ASSATRRRSNRSGASRRISYADPSDEDDMEMELNQARDEDEDEQAKDIENGSD
             LSELSADSNMLEEPEDDDPSEKEDQDPDDKQNGDDDAQAPSPVPVASKAVPGRAMK
             KATLPTRRSARR"
BASE COUNT   1222 a 1320 c 1249 g 1094 t
ORIGIN
alignment_scores
  Quality: 635.00      Length: 1555
  Ratio: 0.772        Gaps: 58
  Percent Similarity: 52.926  Percent Identity: 20.514
alignment_block
  US-09-512-581-2 x EMBIMD ..
  Align seg 1/1 to: EMBIMD from: 1 to: 4885
30  GluMetValArgArgLeuLysMetValVallLysThrPheMetAspMetAs 46
   ::::::::::::::::::::: ::::::::::: :::::::::::
271 GATTGTGTCAGCGCTGCAACGCTAGCGCAGGAGTTCGCCGACGTCGA 320
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
46  pGluAspSerGluGluLysGluLeuTyLysLeuAsnLeuAlaLeuHisL 63
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
321 ACAGGAG...GAGATTGAGAAGAGTCTCTAAGGAAGGCTCCGAGGAGC 367
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
63  euAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeu 79
   ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

```

   ::::::::::: ::::::::::: ::::::::::: :::::::::::
368 TCGCCACGGCTCAACTGCTGGCATAAGGATAGGGGCTTCGGGCCCTGG 417
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
80  ValAlaCysCysLeuAlaAspIlePheArgIleTyLysAlaProGluAlaPr 96
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
418 GCTACTTGCTGCATGTGGATGTGTGGCGCTCTGTGGCTGACGCGCC 467
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
96  oTyThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgG 113
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
468 TTTTACGGCG...AATCAACTAAAGATATTTTACTTGCATTGTGTCGT 514
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
113 In...LeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTy 128
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
515 CGATCATTCGCCGCTAGGGGATCCCTCAATCCTTATATATGCTCCACAC 564
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
129 PheTyLysLeuGluAsnIleAlaTrpValLysSerTyLysAsnIleCysPh 145
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
565 ATTTACGCTTTGAATTCGTCGGGAGGTCAAAAGTATTTGTTCTCATGAC 614
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
145 eGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyLysArgThrLeu 162
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
615 GGATCTAGTACCAGGACACATGATGCTCCACTGTTTATAGCTGCT 664
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
162 heSerValIleAsnAsnGlyHisAsn.....ProAlaHisLysAsnLeuAsn 170
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
665 TTGACATTGTCGAGGCTCAGCTAAAGCCTCGACCGGGAACCAAGTTGCC 714
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
171 GlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCysG 187
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
715 AAAACGCTCGAATATGATATGACCCGCTGCTAGTAGCAGATTATTGACGA 764
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
187 uGlyAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeu 204
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
765 GTCCGAGTCTCTCGCGCTGATGTTGTGATGATGATGATGATGATGATG 814
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
204 al.....ProAlaHisLysAsnLeuAsn 211
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
815 TGCCTGTGTCGCTCTCGCGCTGCTGATGATGATGATGATGATGATGATG 864
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
212 LysGln..... 213
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
865 CCCGAGACTCAGTGGAGGAGAAACAAGAGACGCTGCTACTAAAGGACTA 914
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
214 .....AlaTyLysLeuAlaLysAlaLeuLysArgThrAlaGlnA 228
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
915 CCCGCTGCTGACAACTGCGCAAGCAATCTGCGAGGCTTCCCGCGGAGA 964
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
228 laIleGluProTyLysIleThrThrPhePheAsnGlnValLeuMet..... 242
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
965 GGATGACTAGTCATATCAGCAGTACTTTAAACAAGCTCATAATTGACGCC 1014
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
243 .....LeuGlyLysThrSe 247
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
1015 TCAGCCACGGGAACCAAAATGGGCCCTCAAAAGCAAGCTCGCAGAACTAA 1064
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
247 rIleSerAspLeuSerGluHisValPheAsp..... 257
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
1065 CCTCGATGACTCGGACGAGGAGGAGGAGACATCAGGAATTGAGTAAAG 1114
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
258 .....LeuIleLeuGluLeuTyLysAsnIleAspSerHisLeuLeuLeu 271
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
1115 CGCATCGATTGATCCGTGAGCTGAGAGAGCATGCCCGACGCTTTTGCAG 1164
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
272 SerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluG 288
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
1165 AATGTTATCCCCCAGCTAGAGAGAGAACTGTCGCGCAATCGGTGCTATT 1214
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
288 uArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspS 305
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
1215 CGCGTTG.....TTAGCAACCCAGACCATCGCGGACCTGACAT 1252
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
305 er.GluLeuAlaSerGlnAsnLysPro.....LeuTrp.GlnCyst 318
   ::::::::::: ::::::::::: ::::::::::: :::::::::::

```


1962 CGGCTGAGGAGTCCGACTACGGCCGAGGACGAGCTACTCTGATCCGA 2011
 350 AspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaI 366
 2012 GGGTTGGAGAGAGCTCTCT.....GACAGTATGATAAAGGT 2049
 366 eArgHisAspValIleValSerIleValThrAlaAlaLysAspIleL 383
 2050 CAGACTTCAGCTGTTAAAGCTGTGAGAGCTTCAAGTTCAGGATATCA 2099
 383 euleu.....ValAsnAspHisLeuLeu 390
 2100 TTGCCAAGCTAGGACCTAATGGAGGAGTAGGCAAGGATGGTTCGTCTC 2149
 391 AsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysG 407
 2150 AACACGTTGGCCGATCGCTGTCCGCGAGAGAAACCCGCGTTAGAGTGC 2199
 407 uAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerA 424
 2200 CGCCATGTCACCTGTGGCTAAACTCTGGGCTGTCCGGACTGGAGATGC 2249
 424 laAlaGlyLysAspAla....AlaLysGlnIleAlaTrpIleLysAspLys 439
 2250 TTGCTGAAATGAAGCGGTAAACAGCAGCGTTGCTGGCATCCCGTCACGC 2299
 440 LeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValG 456
 2300 ATCTACAAACGCTTTTACGCCAATGACCTGGAGTGAACCTCTTGATGGA 2349
 456 uArgIlePheAlaGlnTyrMetValPro..... 465
 2350 TCGGCTCATCTACGTCTCTGGTGTCTCTGGGGTTTCCGCCACCCCAAGA 2399
 465 465
 2400 AGCGACACGAGGAATTCCAAAGCCCAACGCAACTCCCAATCGCAATCGCG 2449
 466HisAsnLeuGluThrThrGluArgMetLysCy 476
 2450 AATCGCGCATCCATCGACCATGATGCTATCAGGGCTGAGAGG..... 2491
 476 sLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuA 493
 2492 .ATCTCTCTCTGCGCGCTCCTCAGCAGCGCGCGCAAGAGCATCTCT 2540
 493 snGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509
 2541 TCGGCATGCACTCTAGACGGCTCAGTTCCGAAAGATTTTGGAAACCTAC 2590
 510 LeuAspLeu.....IleLysGlnProLysThrAs 519
 2591 CTCGACACGATGCGATCGGTACACGGTGGTGTATGGAAGCAACGCCGA 2640
 519 pAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnL 536
 2641 TAAATCACCAGCAACCTGACAAACCGCGCATATCATCGCAGATTCC 2690
 536 euProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnVal 552
 2691 TTCCAGAACACGTTAAGAGCAAGACGACCTGATCAAGTTTGGCAAGATT 2740
 553 LeuGluAspAspGluLysIleArgLysGlnLeuValLeuValSerPr 569
 2741CACGATCGCCCAATTAACCTGATCAAAATATGTCATTGGCCA 2784
 569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysL 586
 2785 GGAGAAATGACTTTAAGCGGTTTATAGGCTTTGAAGAGCTTATCAAGC 2834
 586 ysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLys 602
 2835 GCTGTATGGCCAGCAAGACCCCTCA.....GTTATCGACACTCTGCTT 2878

603 PheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSe 619
 2879 CTTTGGCTTACCGTCCGGTTGCTTCTTCAACCCGACCATCTGTC 2928
 619 rAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspG 636
 2929 TAGCATCATGAATACTCCAATCAGACAAAGGAGCGGCTGGGATCG... 2974
 636 luAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeu 652
 2975ATTGGCAGCAGATC 2989
 653 LeuLysValLeuSerPheThrHisProIleSerPhe.....HisSerAl 667
 2990 TTGAACGAGATCTCTCAACGCAACCCGACCTATTCAAAACTCATATCGG 3039
 667 aGluThrPheGluSerLeuLeu.....AlaCysLeuLysMetA 680
 3040 ACAGCTTTGCAAGGATCTTGTGACCAGGCCCTTACTGCCACAAGCCGA 3089
 680 spAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly 696
 3090 ACGATCCCATAGTAGCGAA...ACGCTCAAGGCTTGTCTCCACGTATGCC 3136
 697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProVa 713
 3137 CGGAAGTTCCTTAGGATGTCGCCATCGGATCGGAAGTTCGTGCAGACCAT 3186
 713 lLeuHisHisLysSerLysGlyProProArgGlnAlaLysTyrAlaI 730
 3187 GATCAACTACGCGCTTTATGCCAACCCGTCAGGCTTCGAAACATGCTG 3236
 730 leHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAla... 745
 3237 TCAACATTTGTTTGTAAAGCAGGATGACAAGAGCATGGTCACCTGCCACT 3286
 746 GlnIlePheGluProLeuHisLys.....SerLeuAspProSerAsnLe 760
 3287 GACATTCCTCCAGCGGATTTGAAGGGTTGGTCGTATGGTCGAGTAAC... 3334
 760 uGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuA 777
 3335TTCTTAAACAAGCTTACCGCTGTGAGCCAGTTGGAGCTTCTTG 3377
 777 laProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIle 793
 3378 CACCC...AAGGTTACTGAGGATGCCAGTATGAGATCCTCAATCTCACCC 3424
 794 ValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLy 810
 3425 TTTAAGGAGATTTCTTCCAGGTCGCCAGCCAGCCCAAGACACGCCACCC 3474
 810 sLeuTrpValProAspGluGluValSerProGluThrMetValLysIleG 827
 3475 CGATTGGTCAATGGCCCTGATATGGAGGAGGAATCCAGGCCAAGTCC 3524
 827 lnAlaIleLysMetMet.....ValArgTrpLeuLeuGlyMetLys 840
 3525 TCTCAGTGAATAACTAGTCAATAGGGAAGGAGCATCGAGGGTATCGAG 3574
 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrI 857
 3575 GAGCAAGGAGGAAGGCC...AGTTCTGTTTGAAGGGTGTCTCAGGAAGAT 3621
 857 eLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProA 874
 3622 CATCAAGGAGAGAGGCGAGATGTCGAGGAGAGGAAGAAACGCCAAGCATC 3671
 874 spMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
 3672 ACAAGACGAGGCTGAGGCTTCTGCGCGCACAGTTGATGCTGAAGCTCTGC 3721

SRDDWSPGAAPESERRQSSRRRAANKAKGPTYTERDSDDDDEMLEGVAEWEYLDDEDD
 EDKEDDESGEAPRRTAAKASATRRSRREDKSEEEVNEEPKQEEBEPADPE
 QEEAPTAARAKRTATAGKKAASLPVRNAAASRRSGRKAASAAAAAKDKDDSD
 LSEUSDVEMWDEPEAEAEAEAEVEAEVEEEOEAPPARANGRKAKAAPAK
 AKAAAPGKAKAADASAPARRSGRSTRGKQOPEDEDDANDDEE"
 894...955

intron
 /gene="SP076"
 /number=1

exon
 956...4405
 /gene="SP076"
 /number=2

intron
 4406...4478
 /gene="SP076"
 /number=2

exon
 4479...5649
 /gene="SP076"
 /number=3

polyA_site
 5649
 /gene="SP076"

/evidence-experimental
 5887

polyA_site
 /evidence-experimental

BASE COUNT 1642 a 1734 c 1718 g 1317 t
 ORIGIN

alignment_scores:

Quality: 697.50 Length: 1652
 Ratio: 0.819 Gaps: 60
 Percent Similarity: 51.574 Percent Identity: 21.247

alignment_block:

US-09-512-581-2 x SMAJ9934 ..

Align seg 1/1 to: SMAJ9934 from: 1 to: 6411

13 IleThrTyrProGlyVallysGluIleSerAspLysIleSerLysGlu 29
 639 TTGACATGGCCCGCGC.....AGCCCATTTTCGTTGA 673
 29 uGluMetValArgLeuLysMetValVallysThrPheMetAspMeta 46
 674 TACACTGCTTAAGAGACTCGACGCCCTCTCCAAAGGAGCTGTCGGATATGG 723
 46 spClnAspSerGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHis 62
 724 ACCAGGAGACTGTGGACACCGCATTCGCTG...GTCAAGGTCGCTAAGGAC 770
 63 LeuAlaSerAspPheLeuLysHisProGlyLysAspValArgLeuLe 79
 771 GTGCGCTCTCACAGCTTATCCAAACAAGGACAAAGGGCGTACGACATA 820
 79 uValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAla 96
 821 CACAGCATGCTGCATGTGCATATCTTAAGACTATCGCCCGCGATGCGC 870
 96 roTyrThrSerProAspLysLeuLys..... 104
 871 CCTTTACC...CCATCACAACCTCAAGGTTTGAGAGCCAAAAGATGCCAT 917
 105AspIlePheMetPheIleThrArg..Gln..... 113
 918 TTTATGGACCATCTGCTGATAGCTGGGACCGGACACAGGACATCTTTAAT 967
 114LeuLysGlyLeuGluAspThrLysSerProGlu 124
 968 CTCAGCATAACTCCATCATTCGCCGCTTTTCGATCCCTCGATCCCTA 1017
 124 nPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaIleValLysSerT 141
 1018 CAACAACACGACCAAGTATGTGTACGGTCTTTTTCGCGAGATTAAGAGTA 1067
 141 yrasnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeu 157

1068 TCGTTCGTCTACTCGATGTCGAAGCGACGCGCCTTGTTACTCAAGCTT 1117
 158 TyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLys.... 172
 1118 TTTACCACAATTTTCGACGGGTCTCAGGTCTCAAAATCTCCAAGGAGA 1167
 173ValHisMetHisMetValAspLeuMetSerSerI 184
 1168 GCAGGTTGGAAGGACGTCGAGTTTATGCAAGAGATGTTGGGAGCGC 1217
 184 leIleCysGluGlyaspThrValSerGlnGluLeuLeuAspThrValLeu 200
 1218 TCATTGACGACTCCGTGACCCCTTCCAGGAAGGTGCTCGATGTCATCATG 1267
 201 ValAsn...LeuValProAlaHisLysAsnLeuAsnLysGlnAlaIleTyrAs 216
 1268 GCCAGTTCTCGAGACTGCTGCTCGGGACTAGGCAAGAACGCCAGGA 1317
 216 p.....LeuAlaLysAlaLeuLeuLysArgThrA 226
 1318 TCATGTGCCAATCGACGACGACGCAACACTGCTCTGAAGGAAGAGC 1367
 226 laGlnAlaIleGlu..... 230
 1368 CTGAGGCTTACCAGATGTCAGGAATCTATGCCAGACTTATGACGACAAG 1417
 231ProTyrIleThrThrPheAsnGlnValLeuMet..... 242
 1418 ATGGCCCGTTTCGCCAGTCAATATTTTAGTGCAGTGTGTCGACCGGAC 1467
 243LeuGlyLys..... 245
 1468 AGGTTTGTGGGAAGCGCAAGTGTAGCAGACGAAAGATGACGAAAACG 1517
 246ThrSerIleSerAspLeuSerGlu... 253
 1518 ATGAGGCTGATGCCCAACAGGCCCTTCGAGTCCGACTTGAAGGAATTG 1567
 254HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspse 267
 1568 AGGAAGCGCATGTG.....CTTATTCGGAGATCTGGAGGCTCCCC 1611
 267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSera 284
 1612 CATGATCTCTCAAAAGCTTATCCCGCAGGTGATGCAGAGCTCTCGGCAG 1661
 284 snAspAsnGluGluArgLeuGlnValValLysLeuAlaLysMet... 299
 1662 ACAATGTCACCTGGCAGACAGATGGCAACGGAGACCCTCGGACATGATT 1711
 300PheGlyAlaLys..... 303
 1712 TCAGGATTTGGCGCGCAGGTCGGCTCCACTGCCATCTTGTATCCAGC 1761
 303 303
 1762 TGCTTACCGCCATTGAGCCTGGAGAGGAAGATAGAGCGGAACCCCCAG 1811
 304AspSerGluLeuAlaSerGln 310
 1812 TGACCAACATCTCACCACGCGCTTGTCTCCATCTCTTCGCGACAGA 1861
 311 AsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisVa 327
 1862 CATGACGACCGCTCCCAACATTTCTGAGTAGAAGAATGATGAAGGCTCC 1911
 327 lProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMet.... 342
 1912 CTCATTTCGTGCGGATGGACCGCAGTAGGCCACATTTCTCTTACTT 1961
 343AsnHisProAspLeuAlaLys 349
 349

203 AACCAGAGATGCCAGCTCTCCAGATGAATCCAGACAAATCAAAACTGT 252
1066 yrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThr 1082
253 ATACGATGTGATGGCTCTCTGTGTATATAAATAGTAAAGTCTTG 302
1083 TyrSerLeuGluSerProLysAspProValLeuProAlaArgPheTh 1099
303 TGCATCCAGATTCACCAAGAGCCAGTCTCCCAATGAAATTTTAC 352
1099 rGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluM 1116
353 ACAACCTGAAAGGACTCTGTACAGATAAGAGTTATATTCAGACAGA 402
1116 etLysSerPheThrProGlyLysProLysThrThrAsnValLeuGly 1132
403 CAAGAGTACTTCTGTAAACAGGAAGCAAGCTGCTGGAGTACTAGT 452
1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSe 1149
453 GCAGTAATAAGCTTTATCAGCACCGGGAAGAACCTATGTAGAAC 502
1149 rSerArgMetGluThrValSer.....AsnAlaSerSerSerAsnP 1164
503 CACTGGCACTGAGACTGGAAGCAATATTAATGTAATTCAGAGCTGAAC 552
1164 roSerSerProGlyArgIleLysGlyArg...LeuAspSerSerGluMet 1179
553 CTTCACCGGAATTCGATCAAGGACAGAGTTTCAGAGCGACAGAACT 602
1180 AspHisSerGluAsnGluAsp.....TyrThrMetSerSerProLeuPr 1194
603 GGAGTTAGTGAATAAGAGAACCCCTGTGAGGATATTTCAGTCACACC 652
1194 oGlyLysSerSerAspLysArgAspSerAspLeuValArgSerGluL 1211
653 TGTAAGAATAATTGAC..... 668
1211 euGluLysProArgGlyArgLysLysThrProValThrGluGlnGlu 1227
669 686
1228 LysLeuGlyMetAspAspLeuThrLysLeuValGlnGlnLysProLy 1244
687 ...ATTAATTCATCAGCTACC.....CAGGCAACAT 718
1244 sGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspG 1261
719 CAGCAGTACCCGAGGAAAGAAAGAACAGTAAACAGCAGCTGGTGCAGAGA 768
1261 lu...GlnGlnTrpProGluLysArgLeuLysGluAspIleLeuGlu 1276
769 ATATCCAAACAAAAACAGATGAGAAA..... 794
1277 AsnGluAspGluGlnAsnSerPro...ProLysLysGlyLysArgGlyAr 1292
795 ...GTAGATGAATCGGACCTCCCGCCCTCCCAACCCAGGAGAGACG 841
1292 gProProLysProLeuGly...GlyGlyThrProLysGluGluProThrm 1308
842 TCGACCCCAAGTCTGAATCTCAGGCAATGCTACCAAAAAATGATCATAA 891
1308 etLysThrSerLysGlySerLysLys...SerGlyProProAlap 1324
892 ATAAACCTATTAAACAGGAAGGAGAGAGCTGCAGTGGGTCAGGAGAGC 941
1324 ro.....GluGluGluGluGluGlu 1331
942 CTGGGGTGGAGAGAGGTAATGCCAACACCACTGCAAGATT 991
1332 ArgGlnSerGlyAsnThrGluGlnLys.....SerLysSerLy 1344
992 AGCCAAAAAGGCAGCAGCAGCAGAGAAACAAATTTGACTTACAAAGGTAAA 1041

1344 sGlnHis.....ArgValSerArgAlaGlnGlnArgAlaGluSerp 1359
1042 AATGATTTCGAAGGAGAAATGAAGCCAAACAGAAG...CAGGCTC 1088
1359 roGluSerSer 1362
1089 CAGCTTCTGCA 1099
seq_name: gb_pl4:SMJ9934
seq_documentation_block:
LOCUS SMJ9934 6411 bp DNA PLN 30-JUL-1999
DEFINITION Sordaria macrospora SP076 gene.
ACCESSION AJ009934
VERSION AJ009934.1 GI:5679711
KEYWORDS SP076 gene.
SOURCE Sordaria macrospora.
ORGANISM Sordaria macrospora
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Sordaria.
REFERENCE 1 (bases 1 to 6411)
AUTHORS van Heemst,D., Poggeler,S. and Zickler,D.
TITLE The SP076/PDS5/BIMD gene involved in mitotic sister chromatid
cohesion is also needed for synaptonemal complex formation and
recombination in Sordaria macrospora
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6411)
AUTHORS van Heemst,D.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1998) van Heemst D., Agricultural University,
Department of Genetics, Dreijenlaan 2, 6703HA Wageningen, THE
NETHERLANDS
FEATURES
Location/Qualifiers
1..6411
/organism="Sordaria macrospora"
/db_xref="taxon:5147"
mRNA join(138..893,956..4405,4479..5649).
gene /gene="SP076"
138..5649
exon /gene="SP076"
138..893
/gene="SP076"
/number=1
CDS join(543..893,956..4405,4479..5648)
/gene="SP076"
/codon_start=1
/product="Spo76 protein"
/protein_id="CAB51808.1"
/db_xref="GI:5679712"
/translation="MAPRSAPPEEELEEEQEQGLVALEFNEPLTVRRPKPTISVD
TLKKRLDLSKELSDMDQETVDTLSLVKAKDVASHQLQHKDKGVRATACIVDIL
RLCAPDAPTPSQLKIDFNLSINSIIPALFDPSPNNQHKYLVLSFAIKSIVLLD
VEGSEALLKLETTIFDVGSLKSSGKEQGVKDFEFSMOEMGALIDDSVTLPKVV
VIMAFPLRAAPGLKERODHPIDDSOATLLKKEPEAYQVRLNCOYTDKMARFA
SOYSDVIDATGFEAGKNGSDEEDENDEGGDGTGSPESDLKELKAWHLREIWA
AEMILQNIPOVDAELSDNVHLRQMATETLGDMSGICGAAGPPPLPILDPAAYPPLS
LEEDRAEPPVTNLTPLICSISFSQTHSTTFHNELSRKNDRKAPSIRAAWTAVGHIL
TSAGGIGLSREDEATLIRGLGKELSDSEKYLRSKAVATFKFQDIIAKLPGNGV
GKDSVLNTLADRCRDRKPAVRAAMSLLAKLWAGTKATNSNANGSOSANAASIDHA
YNAFYANDLEAVLMDRVIEFLVPLGPPPAKATNSNANGSOSANAASIDHA
IRAEITLLARSILDRPAKKAFFAMOSRRPQFAKILETYLDQCDRYNGVMSNADKIT
SNLNTADYIAQFLPEHVSKTDLIKFAKHBRNNYLLIKYVIGENDFKTYKALKE
LTKRMASKDPSVIDTLLPLLYRSGCLLENRSHLSIMEYSKDKDGLGSLIAHEILNE
ISORNPDLFTKHIGCKDLVDQAPTATPNPDIVAETLKACSTYARKPKDVMDRK
FYQTMVYALYQGPVKASHAVNIIVLCQDDKSMVTATDILQILKLGKSYGSSNFNLFK
LTASOLELLAPKTEDASDEILNLTKEILLQVTRDAKSDPDQWNGADMDEEIOAK
CLSLKLVNRVSIIEGIEEAKKASVWKLKIKKEKGEIVEEKETPKHKHTRLL
AAQLMKLCTQKHDDMLTPSDFNLLALTQDMVEVRHGFVRKKLQKYLADKLSRY
YTIIFLTAPEPSVEFKNRVETWIRSRARFQNLQPVLEAIWRLISLIAHPDYSNE
VEYLDWAFYILNYIVLVGTESNLGLIYKAEVRKQTDGLNPNSDAHLVSLDAQSI
IRKWEKKWFAAFPGKVLGLYTLALQSHSEAAQIAEKSYLPBGLDKLELIRA
MDRKKRKRKSTSAATGTNNRDGHKSKVKSSGDRDGBREAKSSRKIKATAKSKTAPKRR

* 53814 53913: gap of 100 bp
* 53914 54675: contig of 762 bp in length
* 54676 54775: gap of 100 bp
* 54776 55537: contig of 762 bp in length
* 55538 55637: gap of 100 bp
* 55638 56413: contig of 776 bp in length
* 56414 56513: gap of 100 bp
* 56514 57278: contig of 765 bp in length
* 57279 57378: gap of 100 bp
* 57379 58149: contig of 771 bp in length
* 58150 58249: gap of 100 bp
* 58250 59023: contig of 774 bp in length
* 59024 59123: gap of 100 bp
* 59124 59906: contig of 783 bp in length
* 59907 60006: gap of 100 bp
* 60007 60792: contig of 786 bp in length
* 60793 60892: gap of 100 bp
* 60893 61663: contig of 771 bp in length
* 61664 61763: gap of 100 bp
* 61764 62539: contig of 776 bp in length
* 62540 62639: gap of 100 bp
* 62640 63422: contig of 783 bp in length

alignment_scores:
Quality: 768.00 Length: 147
Ratio: 5.224 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-512-581-2 x AC016449/rev ..

Align seg 1/1 to reverse of: AC016449 from: 1 to: 72157

1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluG1 1225
39685 TCTGAATTGGAGAGCCATAGAGCGAGGAAAAACCCCGTCACAGACCA 39636
1225 nGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluInL 1242
39635 GGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGA 39586
1242 yProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258
39585 AACCTTAAGGACAGTCAGCGAAGTCGAAAAAGAGCCATACGGCTTCAGAA 39536
1259 SerAspGluGlnGlnTrpProGluLysArgLeuLysGluAspIleLe 1275
39535 TCTGATGAACACAGCTGGCGCTCAGGAAAAAGAGGCTCAAGAAGATATATT 39486
1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyA 1292
39485 AGAAATGAAGATGAACAGAAATAGTCCGCCAAAAAAGGGTAAGAAGGCC 39436
1292 rGProProLysProLeuGlyGlyThrProLysGluGluProThrMet 1308
39435 GACCACCAAAACCTTGGTGAGGTACACCAAAAGAGAGCCACCAATG 39386
1309 LysThrSerLysGlySerLysLysLysSerGlyProProAlaProG1 1325
39385 AAAACTTCTAAAAAAGGAAGCAAAAAAATCTGGACCTCCAGCACCAGA 39336
1325 uGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerL 1342
39335 GGAGGAGGAAGAAGAAGCAAAAGTGAATACGGAACCAAGAAGTCCA 39286
1342 ySerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355
39285 AAGGACACAGCAGCCAGTGTCTAAGAGAGACACACAGCAGAGG 39245

seq_name: gb_pr5:AK021757

seq_documentation_block:

LOCUS AK021757 2212 bp mRNA PRI 29-SEP-2000

DEFINITION

Homo sapiens cDNA FLJ11695 fis, clone HEMBA1005019, highly similar to Homo sapiens mRNA for KIAA0648 protein.

ACCESSION

AK021757

VERSION

AK021757.1 GI:10433003

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to mRNA, clone_lib:HEMBAL clone:HEMBAL005019.

ORGANISM

Homo sapiens

REFERENCE

1 (sites)

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Oshima, A.

NEDO human cDNA sequencing project

Unpublished (2000)

2 (bases 1 to 2212)

Isogai, T. and Otsuki, T.

Direct Submission

Submitted (23-AUG-2000)

Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

University of Tokyo.

Location/Qualifiers

1. .2212

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HEMBAL005019"

/clone_lib="HEMBAL"

/dev_stage="embryo, 10 weeks"

/tissue_type="whole embryo, mainly head"

/note="cloning vector: pME18SFL3"

BASE COUNT 751 a 385 c 430 g 646 t

ORIGIN

alignment_scores:

Quality: 706.50 Length: 404

Ratio: 2.656 Gaps: 15

Percent Similarity: 65.842 Percent Identity: 42.822

alignment_block:

US-09-512-581-2 x AK021757 ..

Align seg 1/1 to: AK021757 from: 1 to: 2212

983 LeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLe 999

3 TTATTATCACCTGTCCTGAATATGCTAGTTCCATCATGATTCACCTGCT 52

999 uAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysA 1016

53 ACCCATCATCCAGATTTTACAAAGATCACAAGATGTTGATCAGCTTCGTC 102

1016 spValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032

103 ATATCAAAAGAGTGCCTATGTTTCATGCTTGAAGTTTAAATGACAAAGAT 152

1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysCl 1049

153 GAAAAACAATAGCCATGCCCTTTATGAAGAGATGGCAGAGAACATCAAGTT 202

1049 nThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuT 1066

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1189

Center clone name: 45_L14

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 858 1635: contig of 778 bp in length
* 1636 1735: gap of 100 bp
* 1736 2490: contig of 755 bp in length
* 2491 2590: gap of 100 bp
* 2591 3343: contig of 753 bp in length
* 3344 3443: gap of 100 bp
* 3444 4224: contig of 781 bp in length
* 4225 4324: gap of 100 bp
* 4325 5088: contig of 764 bp in length
* 5089 5188: gap of 100 bp
* 5189 5968: contig of 780 bp in length
* 5969 6068: gap of 100 bp
* 6069 6844: contig of 776 bp in length
* 6845 6944: gap of 100 bp
* 6945 7701: contig of 757 bp in length
* 7702 7801: gap of 100 bp
* 7802 8573: contig of 772 bp in length
* 8574 8673: gap of 100 bp
* 8674 9430: contig of 757 bp in length
* 9431 9530: gap of 100 bp
* 9531 10295: contig of 765 bp in length
* 10296 10395: gap of 100 bp
* 10396 11139: contig of 744 bp in length
* 11140 11239: gap of 100 bp
* 11240 11988: contig of 749 bp in length
* 11989 12088: gap of 100 bp
* 12089 12874: contig of 786 bp in length
* 12875 12974: gap of 100 bp
* 12975 13750: contig of 776 bp in length
* 13751 13850: gap of 100 bp
* 13851 14638: contig of 788 bp in length
* 14639 14738: gap of 100 bp
* 14739 15520: contig of 782 bp in length
* 15521 15620: gap of 100 bp
* 15621 16381: contig of 761 bp in length
* 16382 16481: gap of 100 bp
* 16482 17248: contig of 767 bp in length
* 17249 17348: gap of 100 bp
* 17349 18123: contig of 775 bp in length
* 18124 18223: gap of 100 bp
* 18224 18977: contig of 754 bp in length
* 18978 19077: gap of 100 bp
* 19078 19860: contig of 783 bp in length
* 19861 19960: gap of 100 bp
* 19961 20732: contig of 772 bp in length
* 20733 20832: gap of 100 bp
* 20833 21590: contig of 758 bp in length
* 21591 21690: gap of 100 bp

* 21691 22446: contig of 756 bp in length
* 22447 22546: gap of 100 bp
* 22547 23333: contig of 787 bp in length
* 23334 23433: gap of 100 bp
* 23434 24212: contig of 779 bp in length
* 24213 24312: gap of 100 bp
* 24313 25093: contig of 781 bp in length
* 25094 25193: gap of 100 bp
* 25194 25958: contig of 765 bp in length
* 25959 26058: gap of 100 bp
* 26059 26839: contig of 781 bp in length
* 26840 26939: gap of 100 bp
* 26940 27705: contig of 766 bp in length
* 27706 27805: gap of 100 bp
* 27806 28568: contig of 763 bp in length
* 28569 28668: gap of 100 bp
* 28669 29419: contig of 751 bp in length
* 29420 29519: gap of 100 bp
* 29520 30311: contig of 792 bp in length
* 30312 30411: gap of 100 bp
* 30412 31202: contig of 791 bp in length
* 31203 31302: gap of 100 bp
* 31303 32082: contig of 780 bp in length
* 32083 32182: gap of 100 bp
* 32183 32934: contig of 752 bp in length
* 32935 33034: gap of 100 bp
* 33035 33796: contig of 762 bp in length
* 33797 33896: gap of 100 bp
* 33897 34683: contig of 787 bp in length
* 34684 34783: gap of 100 bp
* 34784 35552: contig of 769 bp in length
* 35553 35652: gap of 100 bp
* 35653 36406: contig of 754 bp in length
* 36407 36506: gap of 100 bp
* 36507 37294: contig of 788 bp in length
* 37295 37394: gap of 100 bp
* 37395 38159: contig of 765 bp in length
* 38160 38259: gap of 100 bp
* 38260 39017: contig of 758 bp in length
* 39018 39117: gap of 100 bp
* 39118 39890: contig of 773 bp in length
* 39891 39990: gap of 100 bp
* 39991 40774: contig of 784 bp in length
* 40775 40874: gap of 100 bp
* 40875 41643: contig of 769 bp in length
* 41644 41743: gap of 100 bp
* 41744 42521: contig of 778 bp in length
* 42522 42621: gap of 100 bp
* 42622 43393: contig of 772 bp in length
* 43394 43493: gap of 100 bp
* 43494 44268: contig of 775 bp in length
* 44269 44368: gap of 100 bp
* 44369 45149: contig of 781 bp in length
* 45150 45249: gap of 100 bp
* 45250 46024: contig of 775 bp in length
* 46025 46124: gap of 100 bp
* 46125 46868: contig of 744 bp in length
* 46869 46968: gap of 100 bp
* 46969 47720: contig of 752 bp in length
* 47721 47820: gap of 100 bp
* 47821 48596: contig of 776 bp in length
* 48597 48696: gap of 100 bp
* 48697 49462: contig of 766 bp in length
* 49463 49562: gap of 100 bp
* 49563 50335: contig of 773 bp in length
* 50336 50435: gap of 100 bp
* 50436 51207: contig of 772 bp in length
* 51208 51307: gap of 100 bp
* 51308 52078: contig of 771 bp in length
* 52079 52178: gap of 100 bp
* 52179 52954: contig of 776 bp in length
* 52955 53054: gap of 100 bp
* 53055 53813: contig of 759 bp in length

fragment_chain:10"
273625..277125
/note="assembly_fragment:04232
fragment_chain:11"
277226..282694
/note="assembly_fragment:02725
fragment_chain:11"
282795..284888
/note="assembly_fragment:04893
fragment_chain:12"
284989..290372
/note="assembly_fragment:01882
fragment_chain:12"
290473..298809
/note="assembly_fragment:00742"
298910..301400
/note="assembly_fragment:00762"
301501..303502
/note="assembly_fragment:01554"
303603..306492
/note="assembly_fragment:02456"
306593..309852
/note="assembly_fragment:03413"
309953..312398
/note="assembly_fragment:03967"
312499..315866
/note="assembly_fragment:04102"
315967..318753
/note="assembly_fragment:04448"
318854..323190
/note="assembly_fragment:04616"
323291..329923
/note="assembly_fragment:04963"
330024..332405
/note="assembly_fragment:06188"
332506..335631
/note="assembly_fragment:06207"
335732..338304
/note="assembly_fragment:06594"
338405..340969
/note="assembly_fragment:06804"
BASE COUNT 100857 a 71110 c 68176 g 96314 t 4512 others
ORIGIN

alignment_scores:
Quality: 769.00 Length: 223
Ratio: 4.272 Gaps: 2
Percent Similarity: 80.717 Percent Identity: 69.058

alignment_block:
US-09-512-581-2 x AL512630/rev ..
Align seg 1/1 to reverse of: AL512630 from: 1 to: 340969

1204 SerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysTh 1220
51114 TCCTCTTTCTCCAGCTTCAGTTGGAGAGCCCTAGAGTCGGAAGAAC 51065
1220 rProValThrGluGluGluLysLeuGlyMetAspLeuThrLysL 1237
51064 ACCTGTACAGACCTTGAAGAGAAATAGGTATGGATGACCTAAGT 51015
1237 euValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGly 1253
51014 TGGTACAGACAGAAACCTTAAGAGCAGTCAGCAGAGCGCGGAAGAGGC 50965
1254 HisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLe 1270
50964 CGTACAGCTTCAGCTTCAGCAGCAGCAGTCGCTGAGGAGAGAGCA 50915
1270 uLysGluAspIleLeuGluAsnGluAspGluClnAsnSerProProLysL 1287

50914 CAAGAGGAGGCTCTGGAAATGAGGATGAGCAGAACAGCCACCACAAAA 50865
1287 ySGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrProLys 1303
50864 AGGGCAAGAGGAGGAGCCACCAAAACCTCTTGGTGGGGGACATCGAAG 50815
1304 GluGluProThrMetLysThrSerLysLysGlySerLysLysSerG1 1320
50814 GAAGAGCCCAACAATGAAACATCCCAAGAAAGGAAACAAAGAAAAA 50765
1320 yProProAlaProGluGluGluGluGluGluGluGluGluGluGluGlu 1337
50764 ACCTCTCTAGTAGCAGCAGATGAAGAAGAAAGAAAGAAATTTGGAACA 50715
1337 hrGluGlnLysSerLysGlnHisArgValSerArgArgAlaGln 1353
50714 CAGAACATAAGTCAAAAAGCAAAACAGCAGCCGACATCAAGAGAGCA 50665
1354 ..Gln..... 1354
50664 CAGAGGTGAGTGTGTGCAGTGCTAAAGTGGACTTCGGTGCAGTAAATGTG 50615
1354 1354
50614 CTTACCAATAATGTAATCATATTTGATGCTATCCACATTTGGGTCTTCC 50565
1355ArgAlaGluSerProGluSerSerAlaIleGluSerThrGlnS 1369
50564 CCAAGCAGCAGCAGAAATCTCTGAAACAAGTCAGTTGAATCCACACAGT 50515
1369 erThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSer 1385
50514 CCACACCACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 50465
1386 GlnProLysLysAsnVal 1391
50464 CAACCCCAAAAAAATG 50447
seq_name: gb_htg5:AC016449
seq_documentation_block:
LOCUS AC016449 72157 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-45L14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016449
VERSION AC016449.2 GI:9104517
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 72157)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-45L14
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 72157)
REFERENCE 1 (bases 1 to 72157)
AUTHORS Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Collings,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galegani,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,A., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 13, 2000 this sequence version replaced gi:6479175.


```

1355 .....ArgAlaGluSerProGluSerSe 1362
29032 TATCCACATTGGGCTTCCCAACGACGAGCATCTCCGTAATCTAG 28983
1362 rAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProS 1379
28982 TGCAATTGAATCCACACATCTCCACACACAGAAAGGACGAGGAAGACCAT 28933
1379 erlystThrProSerProSerGlnProLysLysAsnVal 1391
28932 CAAAACGCCATCACCATCACCAACCAAAAAAATGTG 28895
seq_name: gb_htg22:AL512630

seq_documentation_block:
LOCUS   AL512630 340969 bp DNA HTG 21-MAR-2001
DEFINITION Mus musculus chromosome 5 clone RP21-583E8, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL512630
VERSION   AL512630.2 GI:13443466
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE    house mouse.
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 340969)
AUTHORS   Sims, S.
TITLE     Direct Submission
JOURNAL   Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequests@sanger.ac.uk
COMMENT   On Mar 24, 2001 this sequence version replaced gi:12193249.
            ----- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouse@har.mrc.ac.uk
            ----- Project Information
            Center project name: dM583E8
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 323281 bases at least Q40
            Consensus quality: 329435 bases at least Q30
            Consensus quality: 333234 bases at least Q20
            Insert size: 336469; sum-of-contigs
            Insert size: 189900; 1.3% error; agarose-fp
            Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality
            coverage: 11.68x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

FEATURES             Location/Qualifiers
     source           1..340969
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP21-583E8"
                     /clone_lib="RPCI-21"
     misc_feature     1..10854
                     /note="assembly_fragment:06116
                     fragment_chain:1"
     misc_feature     10955..16884
                     /note="assembly_fragment:05022
                     fragment_chain:1"
     misc_feature     16985..22344
                     /note="assembly_fragment:04089
                     fragment_chain:1"
     misc_feature     22445..36631
                     /note="assembly_fragment:00629

```

```

/note="assembly_fragment:04629
fragment_chain:1"
36732..63941
/note="assembly_fragment:02431
fragment_chain:1"
64042..76959
/note="assembly_fragment:06652
fragment_chain:2"
77060..80361
/note="assembly_fragment:02856
fragment_chain:2"
80462..84581
/note="assembly_fragment:05282
fragment_chain:2"
84682..89169
/note="assembly_fragment:01199
fragment_chain:2"
89270..94454
/note="assembly_fragment:00854
fragment_chain:3"
94555..99344
/note="assembly_fragment:04769
fragment_chain:3"
99445..105364
/note="assembly_fragment:03592
fragment_chain:3"
105465..112857
/note="assembly_fragment:02797
fragment_chain:4"
112958..116353
/note="assembly_fragment:05253
fragment_chain:4"
116454..119029
/note="assembly_fragment:01663
fragment_chain:4"
119130..157992
/note="assembly_fragment:03653
fragment_chain:5"
158093..164601
/note="assembly_fragment:05568
fragment_chain:5"
164702..195901
/note="assembly_fragment:01977
fragment_chain:5"
196002..198807
/note="assembly_fragment:00023
fragment_chain:6"
198908..204604
/note="assembly_fragment:05254
fragment_chain:6"
204705..207838
/note="assembly_fragment:00245
fragment_chain:7"
207939..221335
/note="assembly_fragment:03587
fragment_chain:7"
221436..224577
/note="assembly_fragment:00642
fragment_chain:8"
224678..228248
/note="assembly_fragment:04656
fragment_chain:8"
228349..237923
/note="assembly_fragment:01666
fragment_chain:9"
238024..242134
/note="assembly_fragment:05529
fragment_chain:9"
242235..250341
/note="assembly_fragment:02430
fragment_chain:10"
250442..273524
/note="assembly_fragment:00629

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature


```

* 15233 15912: contig of 680 bp in length
* 15913 16012: gap of 100 bp
* 16013 16731: contig of 719 bp in length
* 16732 16831: gap of 100 bp
* 16832 17506: contig of 675 bp in length
* 17507 17606: gap of 100 bp
* 17607 18339: contig of 733 bp in length
* 18340 18439: gap of 100 bp
* 18440 19109: contig of 670 bp in length
* 19110 19209: gap of 100 bp
* 19210 19900: contig of 691 bp in length
* 19901 20000: gap of 100 bp
* 20001 20693: contig of 693 bp in length
* 20694 20793: gap of 100 bp
* 20794 21464: contig of 671 bp in length
* 21465 21564: gap of 100 bp
* 21565 22245: contig of 681 bp in length
* 22246 22345: gap of 100 bp
* 22346 23049: contig of 704 bp in length
* 23050 23149: gap of 100 bp
* 23150 23857: contig of 708 bp in length
* 23858 23957: gap of 100 bp
* 23958 24685: contig of 728 bp in length
* 24686 24785: gap of 100 bp
* 24786 25499: contig of 714 bp in length
* 25500 25599: gap of 100 bp
* 25600 26302: contig of 703 bp in length
* 26303 26402: gap of 100 bp
* 26403 27108: contig of 706 bp in length
* 27109 27208: gap of 100 bp
* 27209 27910: contig of 702 bp in length
* 27911 28010: gap of 100 bp
* 28011 28720: contig of 710 bp in length
* 28721 28820: gap of 100 bp
* 28821 29484: contig of 664 bp in length
* 29485 29584: gap of 100 bp
* 29585 30328: contig of 744 bp in length
* 30329 30428: gap of 100 bp
* 30429 31111: contig of 683 bp in length
* 31112 31211: gap of 100 bp
* 31212 31931: contig of 720 bp in length
* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36709: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length
* 38331 38430: gap of 100 bp
* 38431 39109: contig of 679 bp in length
* 39110 39209: gap of 100 bp
* 39210 39912: contig of 703 bp in length
* 39913 40012: gap of 100 bp
* 40013 40712: contig of 700 bp in length
* 40713 40812: gap of 100 bp
* 40813 41514: contig of 702 bp in length
* 41515 41614: gap of 100 bp
* 41615 42294: contig of 680 bp in length
* 42295 42394: gap of 100 bp
* 42395 43094: contig of 700 bp in length
* 43095 43194: gap of 100 bp
* 43195 43915: contig of 721 bp in length
* 43916 44015: gap of 100 bp
* 44016 44740: contig of 725 bp in length

```

```

* 44741 44840: gap of 100 bp
* 44841 45533: contig of 693 bp in length
* 45534 45633: gap of 100 bp
* 45634 46330: contig of 697 bp in length
* 46331 46430: gap of 100 bp
* 46431 47137: contig of 707 bp in length
* 47138 47237: gap of 100 bp
* 47238 47927: contig of 690 bp in length
* 47928 48027: gap of 100 bp
* 48028 48709: contig of 682 bp in length
* 48710 48809: gap of 100 bp
* 48810 49509: contig of 700 bp in length
* 49510 49609: gap of 100 bp
* 49610 50331: contig of 722 bp in length
* 50332 50431: gap of 100 bp
* 50432 51135: contig of 704 bp in length
* 51136 51235: gap of 100 bp
* 51236 51981: contig of 746 bp in length
* 51982 52081: gap of 100 bp
* 52082 52783: contig of 702 bp in length
* 52784 52883: gap of 100 bp
* 52884 53580: contig of 697 bp in length
* 53581 53680: gap of 100 bp
* 53681 54398: contig of 718 bp in length.
FEATURES
      Location/Qualifiers
      source
      1. 54398
      /organism="Homo sapiens"
alignment_scores:
      Quality: 819.50      Length: 196
      Ratio: 5.028      Gaps: 1
      Percent Similarity: 83.163      Percent Identity: 83.163
alignment_block:
US-09-512-581-2 x AC068224/rev
Align seg 1/1 to reverse of: AC068224 from: 1 to: 54398
1227 GluLysLeuGlyMetAspLeuThrLysLeuValGlnGluLysPr 1243
29482 GAGAAATTAGGTATGATGACTGAACCTAAGTTGGTACAGGAACAAACC 29433
1243 oLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerA 1260
29432 TAAAGGCAGTCAGCGAAGTCGGAANAGAGGCCATACGGCTTCAGAATCTG 29383
1260 spGluGlnGlnTrpProGluGluLysArgLysGluAspLeuGlu 1276
29382 ATGAACACAGTGGCCTCAGGAAAAGAGGCTCAAAGAGATATATTAGAA 29333
1277 AsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgPr 1293
29332 AATGAAGATGAACAGACATAGTCCGCCAAAAAGGGTAAAGAGGCCGACC 29283
1293 oProLysProLeuGlyGlyThrProLysGluGluProThrMetLysT 1310
29282 ACCAAACCTCTTGTGGAGGTACACCAAAAGAGAGCAACAATGAATA 29233
1310 hrSerLysLysGlySerLysLysLysSerGlyProProAlaProGluGlu 1326
29232 CTTCTAAAAGGAGCAAAAAAATCTGGACCTCCAGCACCAGGAG 29183
1327 GluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSe 1343
29182 GAGGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAGAGTCCAAAG 29133
1343 rLysGlnHisArgValSerArgArgAlaGlnGln..... 1354
29132 CAACAGACCCGAGTGTCAAGGAGAGACAGCAGAGGTAAGCATGTGTAA 29083
1354 ..... 1354
29082 CTCCTAAACTGCATCTGTTTCGTTACTATTATATAATCATATTGTATGC 29033

```



```

:::||||:..... |||||
5333 AAAATATTCAGCTGAATCCAGAACTTTGGATGGAGTTTGACTACGTA 5382
1083 r.....SerLeuGluSerProLysAspProValLeuProAlaArgPhep 1098
5383 TCCAAAGCAAAATTTAAATACCTTACGAAATTTTAAAGCAAAATTCCTCTA 5432
1098 heThrGlnProAspLysAsnPheSerAsnThrLysAsnThrLysLeuProPro 1114
5433 TAGATGAGAAAAGCAAAATTTCAAT.....AAATCTTCATACGCGCT 5476
1115 GluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValle 1131
5477 AAAATGAAAGT.....CAGATTGA 5496
1131 uGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrL 1148
5497 ACATGCTATTAAGACACACAGCTCTCATCTTTTGGGAACACAGACTACCA 5546
1148 ySerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPro 1164
5547 AGCAGCAAAATTTGAGCAAAAGCACTCATAGCTCAAGACCGATATAA 5596
1165 SerSerProGlyArgLysGlyArgLeuAspSerSerGluMetAspHl 1181
5597 AAATCGTCTAGAAAGGAAA.....AACGAAAAAAGGAG 5631
1181 sSerGluAsnGluAspThrThrMetSerSerProLeuProGlyLysLys 1198
5632 AAAGCTAAATGACGACCAATCTTAATATTAGGAACGTGCCAAGACGA 5681
1198 erAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysPro 1214
5682 GTCAGAGATTCAAGGATTCGATATAAATTATTCGGAAGCACCGCTCT... 5728
1215 ArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGlyMe 1231
5729 .....TCAAGTGAGGAGATAAGCGAAGAGAGAGGAGGAGAAATTAGCGA 5769
1231 tAspAspLeuThrLysLeu 1237
5770 AGAAGATTTTGATGAATA 5788
seq_name: gb_htg14:AC068224

```

```

seq_documentation_block:
LOCUS AC068224 54398 bp DNA HTG 30-APR-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC068224
VERSION AC068224.1 GI:7671284
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 54398)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,E.,
Boguslavsky,L., Boukhgatter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kan,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

```

Meidrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE JOURNAL COMMENT

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: LI0161

Center clone name: 660_H_19

* NOTE: This record contains 68 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

* 1 685: contig of 685 bp in length
* 686 785: gap of 100 bp
* 786 1513: contig of 728 bp in length
* 1514 1613: gap of 100 bp
* 1614 2354: contig of 741 bp in length
* 2355 2454: gap of 100 bp
* 2455 3184: contig of 730 bp in length
* 3185 3284: gap of 100 bp
* 3285 3959: contig of 675 bp in length
* 3960 4059: gap of 100 bp
* 4060 4746: contig of 687 bp in length
* 4747 4846: gap of 100 bp
* 4847 5564: contig of 718 bp in length
* 5565 5664: gap of 100 bp
* 5665 6360: contig of 696 bp in length
* 6361 6460: gap of 100 bp
* 6461 7132: contig of 672 bp in length
* 7133 7232: gap of 100 bp
* 7233 7937: contig of 705 bp in length
* 7938 8037: gap of 100 bp
* 8038 8749: contig of 712 bp in length
* 8750 8849: gap of 100 bp
* 8850 9522: contig of 673 bp in length
* 9523 9622: gap of 100 bp
* 9623 10356: contig of 734 bp in length
* 10357 10456: gap of 100 bp
* 10457 11181: contig of 725 bp in length
* 11182 11281: gap of 100 bp
* 11282 11975: contig of 694 bp in length
* 11976 12075: gap of 100 bp
* 12076 12785: contig of 710 bp in length
* 12786 13885: gap of 100 bp
* 12886 13557: contig of 672 bp in length
* 13558 13657: gap of 100 bp
* 13658 14347: contig of 690 bp in length
* 14348 14447: gap of 100 bp
* 14448 15132: contig of 685 bp in length
* 15133 15232: gap of 100 bp

```


•
•
•
—
—

•
•
•
•
•
•
•

•
•
•
•
•
•

•

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Schizosaccharomyces.
1 (bases 1 to 9982)
Zimmermann, W., Wambutt, R., Wood, V., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (15-SEP-2000) European Schizosaccharomyces genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and Agowa GmbH, Glienicke Weg 185, 12489 Berlin, Germany

Notes:
Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
during 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid c110 is overlapped at the 5' end by cosmid c140, EMBL entry SPAC140, accession number Al163191, and at the 3' end by pl p14E8, EMBL entry SPAP14E8, accession number Al159180.
Location/Qualifiers
1..9982
/organism="Schizosaccharomyces pombe"
/strain="972h."
/db_xref="taxon:4896"
/chromosome="II"
/clone="cosmid c110"
/map="IIL"
misc_feature
1..278
/gene="SPAC140.05"
/note="nominal overlap with SPAC140 S. pombe chromosome 1"
CDS
1..1480
/partial
/gene="SPAC140.05"
/note="Protein sequence is in conflict with the conceptual translation: SPAC110.01"
/codon_start=2
/label=SPAC110.01
/product="putative 5'-amp-activated protein kinase"
/protein_id="CAC08559.1"
/db_xref="GI:10185124"
/translation="MSTKQNCOTSRMLREVAIKLGNHNPVHKVLDVFKTKHFGI VLDVNGSELDYILARPLEDSVACFLAQLISGVYLHSGVYVHRDLKLENIILDT NRNIIVADGFGATTFGHNTLSEHSLPTEKPDLPSTSCGSPCYAAPLVNCKSGMYAG TQADISGVLYLAGLPLFDDDDPHNPNENVPRLYRICSTPLIPFPFCKSKARD ILKILVPDPAKRMASIMNHAYLRHVSFLFTYNDDEATSLRPSIRLLSOKSLAS VSSVSTISSTSELTPTQITKTLNREKPLHGSINLKHSTRPVPIADGIYASTTL HPIDNAKRLVLLRRPKHKVYKAKVETRESPMKNKPTKKOKHRAFSMFERPTSOAP LSPVESKINLVAPKRVLCNADSRLALACKTDKSPGWSRIPFLFNKLKSKQKAYSS SFNTENSELILQVSPSTFGCGFRENNAYGKKQHRHATVF"
1..1480
/gene="SPAC140.05"
/note="SPAC110.01"
11..730
/misc_feature
/gene="SPAC140.05"

gene
CDS
misc_feature
intron
intron
gene
CDS

/note="Match to PF00069 pkinase, Eukaryotic protein kinase domain Score 247.77"
2183..5800
/gene="SPAC110.02"
2183..5800
/gene="SPAC110.02"
/note="SPAC110.02, len:1206"
/codon_start=1
/label=SPAC110.02
/product="Protein with possible role in chromosome dynamics by similarity to yeast pds5"
/protein_id="CAC08560.1"
/db_xref="GI:10185125"
/translation="MIKQFORNTVPHDNPFTTSEILKRLRDLGELTSLSQDTIDR DSVLPVARSLVNNLHKKDKGIRSYTLCICIVELLRLCAPDPTLSQLEDIFQVILK ILSGLMQEETSYQIIVEILSELNKAVALIVDLPAEELFNLFRFFDLARKGTT KNVEYMLDIINOLINENTIPAAALNIPAAQLISGKGVROTIGSSDSTNGHPAFOLA RNIHDSADRLQRYCOYFSDIIEFSDSDSDMTPEEFIFSHNLVLQWKYATVTL NIIFENEQAEQTSVRLVALETVGLMQDNAIWSDFPRVMSAFGRNDKSVACRI KCIQENALQNSLATSIEINNVOMIQSLADDEKVRVATLTKTIEOLTETFMQF SVQALKMDRLDRKLNRLQAIQTLSQIYNRAYQDLIDGVEYSIQMFWIPSSLE VFYVNDETNAAVICMAELVQLYSSDTQRLNRLFLSIKIFYSEKAMFVILLQRO VKYSELNYYIECKNKNGVMDNDEESTNKLKVIDIISKSSNPLTEATPRKFA ELNDROSYKMLLOFESIKSEYQVYLSIKYLFKVSSETLSTASLECEPFIYRSALFA FNKSNVHEIQLLNEPKYHNFLKPSBALLOHLPLIHPNIYGEVIEVENIIVSGIE DSVKIVKALSOFSKRNKFSIQTTTAEILRLKLIHQEQAKAATIIATITKEFKL DMITNIVENLYNGKLRVLMTQGLFYLTLLEEVKADQVTEFLVKVIQRPPEKYD DTHNDEWCTVEKLDNLTMCVLAIRVLNRLRAAGTEALNIGAPIKLKVLMA DGEISPFKNTPKISRAYLRFTASKVFLKCSIPFYAEHIDFSYVOISLQCQDNFV RNLFTLKQLOLKLKPLISYVPLELFTAVDPEEIKTKASIMIRSOVAFQKTHDFT MEYVATLHLLSHHPDISSESNSLDFIAYIRFYVDTVNNSNPVIFVHLMQRIKQ NYVDIEDNNYIVLSDMAQILQVSONFQWSLTPYKQIKLPIELRIPSPIDEKK RIFNKIFITPKMESQIEHAIRTPVSSFAKQTNKHLKQKTHSSKSKDKSSRRKN EKRRKLEQNPNIRNVPERSSRFQIRINYSAPSSSEIESEEEISEEDFDEIED L"
/join(6249..6296,6460..6528,6959..7420)
/gene="cdc42"
/note="SPAC110.03"
/join(6249..6296,6460..6528,6959..7420)
/gene="cdc42"
/note="SPAC110.03, len:193"
/codon_start=1
/label=cdc42
/product="cell division control protein 42 homolog"
/protein_id="CAC08561.1"
/db_xref="GI:10185126"
/translation="MPTIKCVVGDGAVGKTKLLISYTNKFPSPDYVPTVDNYAVTV MIGDEPTLGLFDAGQEDVRLPGLSVPTDVLVCFVSFTSPASFENVKKWPVTH HCPGVCPLIVGTQIDLRDPSVOOKLAROHQPLTHEQGERLARELGAVYVCEAL TQKGLNWFDEAIVAAIDPPVPHKKKSKCLVL"
/join(6261..6296,6460..6528,6959..7417)
/gene="cdc42"
/note="Match to PF00071 ras, Ras family Score 237.97"
6297..6459
/note="confirmed by mRNA"
6529..6958
/note="confirmed by mRNA"
complement(8139..9982)
/gene="psl1"
/note="psl1: SPAP14E8.01c; SPAC110.04c"
complement(8139..9982)
/partial
/gene="psl1"
/note="SPAC110.04c"
/codon_start=3
/label=psl1
/product="heat shock protein 70-like protein Ssp1p"
/protein_id="CAC08562.1"
/db_xref="GI:10185127"
/translation="QYLNTEAFSNLIQIAIYFTKIKAIABEALIGSDVVISIPAW FTDIQRALLAEANAGLIRLMDNNAALTYGTDLPLEPEKPRVIVDFGHS NYSYIVFEFSRQPHIKSTVCDNLGSRNMDKALIDYFAAEFKYKIDVLSNPKATP RLATAVERLKKVLSANANAPLVEMINDIDASSFIKRSDEELIKPLLERUTPTPIEK

1355ArgAlaGluSerProGluS 1361
 21608 ATGCTATCCACATTTGGGTCTTCCCAAGCAGACAGACTCTCTGAAT 21657
 1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArg 1377
 21658 CTAGTGAATTAATTCACACAGTCCACACCACAGAAAGGACGAGGAAGA 21707
 1378 ProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
 21708 CCATCAAAACGCCATCACCATCAACCAACCAAAAAAATGTG 21749

seq_name: gb_pr5:AL138820

seq_documentation_block:
 LOCUS AL138820 168487 bp DNA PRI 01-NOV-2000
 DEFINITION Human DNA sequence from clone RP11-380B4 on chromosome 13, complete
 sequence.

ACCESSION AL138820
 VERSION AL138820.11 GI:11022364
 KEYWORDS HTG.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (01-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonequery@sanger.ac.uk
 On Oct 25, 2000 this sequence version replaced gi:10715762.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-380B4 is from the library RPEC-11.2 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-380B4.

FEATURES

source

1. .168487
 /location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-380B4"
 /clone_lib="RPEC-11.2"

BASE COUNT 48433 a 33155 c 34624 g 52275 t

ORIGIN

alignment_scores:

Quality: 925.50
 Ratio: 5.057

Length: 214
 Gaps: 1

Percent Similarity: 85.514 Percent Identity: 85.514

alignment_block:

US-09-512-581-2 x AL138820

Align seg 1/1 to: AL138820 from: 1 to: 168487

1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluG1 1225
 |||||
 20594 TCTGAATTTGGAGAACCTTAGAGGCGAGGAAAAAACGCCGTCACAGAACA 20643
 |||||
 1225 nGluGluLysLeuGlyMetAspLeuThrLysLeuValGlnGluGlnL 1242
 |||||
 20644 GGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGA 20693
 |||||
 1242 ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258
 |||||
 20694 AACCTAAAGGCAGTCAGCGAAGTCGAAAGAGAGGCCATACGGCTTCAGAA 20743
 |||||
 1259 SerAspGluGlnGlnTTPProGluGluLysArgLeuLysGluAspIleLe 1275
 |||||
 20744 TCTGATGAACAGCAGTGGCTGAGGAAAAAGAGGCTCAAGAAGATATATT 20793
 |||||
 1275 uGluAsnGluAspGluGlnAsnSerProLysLysGlyLysArgGlyA 1292
 |||||
 20794 AGAAATGAAGATGAACAGAAATAGTCCGCCCAAAAGGTTAAAGAGGCC 20843
 |||||
 1292 rgProProLysProLeuGlyGlyThrProLysGluGluProThrMet 1308
 |||||
 20844 GACCACCAAAACCTTCTGGTGGAGGTACACCAAAAGAGAGGCCAACATG 20893
 |||||
 1309 LysThrSerLysLysGlySerLysLysSerProLysLysSerGlyProAlaProG1 1325
 |||||
 20894 AAAACTTCTAAAGAAAGGAAACAAATAATCTGGACCTCCAGCACCAGA 20943
 |||||
 1325 uGluGluGluGluGluGlnArgGlnSerGlyAsnThrGluGlnLysSerL 1342
 |||||
 20944 GGAGGAGGAGAGAAAGAAACACAAAGTGGAAATACGGAACAGAGTCCA 20993
 |||||
 1342 ysSerLysGlnHisArgValSerArgArgAlaGlnGln..... 1354
 |||||
 20994 AAAGCAACAGCACCAGTGTCTCAAGGAGAGACAGACAGAGTAAGCATGT 21043
 |||||
 1354 1354
 21044 GTAACCTAAACTGCATCTGTTTCTTACTATATATATAATCATTAATTG 21093
 |||||
 1355ArgAlaGluSerProGluS 1361
 |||||
 21094 ATGCTATCCACATTTGGGTCTTCCCAAGCAGACAGAGAAATCTCTGAAT 21143
 |||||
 1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArg 1377
 |||||
 21144 CTAGTGAATTAATTCACACAGTCCACACCACACAGAAAGGACGAGGAAGA 21193
 |||||
 1378 ProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
 |||||
 21194 CCATCAAAACGCCATCACCATCAACCAACCAAAAAAATGTG 21235

seq_name: gb_pl4:SPAC

seq_documentation_block:

LOCUS SPAC 9982 bp DNA PLN 16-SEP-2000
 DEFINITION S.pombe chromosome I cosmid cl10.
 ACCESSION AL441624
 VERSION AL441624.1 GI:10185123
 KEYWORDS 5'-amp-activated protein kinase; cdc42; cell division control
 protein 42 homolog; chromosome dynamics; heat shock protein 70-like
 protein; pss1; sps1.

SOURCE

Schizosaccharomyces pombe
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;


```

|||||
388 TCTGATGAACAGAGTGCCCTGAGGAAAGAGGCTCAAGAAGATATTT 437
1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyA 1292
438 AGAAATGAAGATGAACAGAAATAGTCGCCCAAAAGGGTAAAGAGGCC 487
1292 rgProProLysProLeuGlyGlyThrProLysGlyGluProThrMet 1308
488 GACCACCAAAACCTCTGGTGGAGGTACACCAAAAGAGAGCAACATG 537
1309 LysThrSerLysLysGlySerLysLysLysSerGlyProProLysProL 1325
538 AAAACTTCTAAAAAGGAAGCAAAATAATCTGGACCTCCAGCACCAGA 587
1325 uGluGluGluGluGluGlnSerGlyAsnThrGluGlnLysSerL 1342
588 GGAGGAGGAAGAGAGAAAGACAAAGTGGAATACGGAACAGAAATCCA 637
1342 ySerLysGlnHisArgValSerArgArgAlaGlnGln..... 1354
638 AAGCAAAACAGCACCGAGTGTCAAGGAGACACAGACAGAGTAAAGCATGT 687
1354 ..... 1354
688 GTAACCTCTAAATGCATCTGTTCTGTTACTATATATATAAATCATATTTG 737
1355 ..... ArgAlaGluSerProGluS 1361
738 ATGCTATCCACATTTGGGTCTCTCCCAAGCAGACAGAGATCTCCTGAAT 787
1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArg 1377
788 CTAGTGCATTTGAATCCACAGTCCACACACACAGAAAGGACGAGGAAGA 837
1378 ProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
838 CCATCAAAACGCCATCACCATCACACCAACCAAAAAAATGTG 879

seq_name: gb_pr8:H549J10

seq_documentation_block:
LOCUS HS49J10 137246 bp DNA PRI 22-NOV-1999
DEFINITION Human DNA sequence from PAC 49J10, BRCA2 gene region chromosome
13q12-13 contains ESTs.
ACCESSION Z84572
VERSION Z84572.1 GI:1813974
KEYWORDS 13q12-13.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137246)
Direct Submission
Whiteley, M.
Submitted (09-JAN-1997) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
IMPORTANT:
This sequence is not the entire insert of clone . It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 49J10 is at 1 in this sequence. The true
right end of clone 267P19 is at 43908.
The true left end of clone 179I15 is at 137143.
49J10 is from the human PAC library described in Ioannou A.P. et al

```

```

FEATURES
            source
            Location/Qualifiers
                1..137246
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="13"
                /map="13q12-13"
                /clone="XX-49J10"
                1551..1831
                /note="AluX repeat: matches 281. .1 of consensus;
                incomplete repeat"
                repeat_region
                1838..1871
                /note="17 copies of 2 mer 100 % conserved"
                repeat_region
                1882..2741
                /note="L1PA15 repeat: matches 893. .1 of consensus"
                repeat_region
                2597..2848
                /note="L1 repeat: matches 5389. .5139 of consensus"
                repeat_region
                3149..3405
                /note="MER44A repeat: matches 247. .1 of consensus"
                repeat_region
                3853..3922
                /note="AluSc repeat: matches 188. .120 of consensus;
                incomplete repeat"
                repeat_region
                5620..5810
                /note="MIR repeat: matches 58. .262 of consensus"
                repeat_region
                7829..8133
                /note="AluX repeat: matches 1. .301 of consensus"
                repeat_region
                10957..11242
                /note="AluY repeat: matches 291. .1 of consensus"
                repeat_region
                12344..12405
                /note="L1PA16 repeat: matches 903. .842 of consensus"
                repeat_region
                12381..13119
                /note="L1PA16 repeat: matches 833. .68 of consensus"
                repeat_region
                14428..14540
                /note="HY1 repeat: matches 110. .1 of consensus"
                repeat_region
                14951..15071
                /note="MIR2 repeat: matches 126. .1 of consensus"
                repeat_region
                16589..16885
                /note="AluX repeat: matches 302. .4 of consensus"
                repeat_region
                18067..18397
                /note="L1 repeat: matches 3807. .4127 of consensus"
                repeat_region
                19076..19481
                /note="MLT1A2 repeat: matches 374. .1 of consensus"
                repeat_region
                19516..19799
                /note="AluX repeat: matches 19. .302 of consensus;
                incomplete repeat"
                repeat_region
                19808..20127
                /note="AluY repeat: matches 2. .299 of consensus"
                prim_transcript
                <21359..26899
                /note="match: multiple ESTs; match: H44781 D20082 W20070
                T64867 H45693; match: F12354 N91467 N83765 H45741 H93424"
                repeat_region
                24795..25089
                /note="AluX repeat: matches 294. .1 of consensus"
                repeat_region
                26083..26106
                /note="12 copies of 2 mer 96 % conserved"
                prim_transcript
                complement(27120..27476)
                /note="match: multiple ESTs; match F09972 T73607"
                prim_transcript
                <27633..28101
                /note="match: multiple ESTs; match: 5' EST T51695 clone
                72405; Paired with EST T51543 matching this clone; match
                H91918"
                repeat_region
                28327..28360
                /note="17 copies of 2 mer 88 % conserved"
                prim_transcript
                <28864..>29013
                /note="match: 3' EST T51543 clone 72405; Paired with EST
                T51695 matching this clone"
                repeat_region
                29300..29456
                /note="MIR2 repeat: matches 146. .1 of consensus"
                repeat_region
                30648..30934
                /note="AluX repeat: matches 7. .302 of consensus"
                repeat_region
                33156..33439
                /note="AluX repeat: matches 301. .2 of consensus"
                repeat_region
                34900..35062
                /note="L1MD2 repeat: matches 425. .263 of consensus"

```



```
977 laAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993
99016 GCTCAACGGAGTCACAAATCATTTACATACACTACTGCTGCTGCTGCT 58967
994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAs 1010
58966 TTGCGTATTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 58917
1010 pIleGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluI 1027
58916 CTACGTACAGCTGCGCAAGATCGAGAGTGCCTGCTGCTGCTGCTGCTG 58867
1027 leuLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMet 1043
58866 CGCTGATGCGCAACAGAGAAAGCTTTGTCATAGCTTCTACAGCAGCTG 58817
1044 ValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaL 1060
58816 CTGCAGCTGATAAGCATCGGAGTTCAGTCTGCGG....TCGGACAAGCG 58770
1060 sMetAsn..... 1062
58769 CGACAACATATGTATGTCACCTCAATCATTTTCCAGGCTTTCCCTCTT 58720
1063 .....GluLysLeuTyrThrValCys 1069
58719 TTATATTTATACATATATTTTTCGGTTATCAGAAATGTGGCGCTCTGC 58670
1070 AspValAlaMetAsnIleLeuMetSerLys..... 1079
58669 GATCTTGCCATGTACATATTCAGCTCAAGTTCAGTCCATTTGTGTGGCAA 58620
1080 SerThrTyrSerLeuGluSerProLysAspProValLeuProAlaA 1096
58619 CAGGACACCTTTTCCATCGCGCTG.....GCTTCCAGAGAA 58582
1096 rgPhePheThrGlnProAsp...LysAsnPheSerAsnThrLysAsnTyr 1111
58581 TGTATTATAAAGAGCTGCGCTGCGAATTTTCCAAACAATGACGCTCTAT 58532
1112 LeuProGluMetLysSerPhePheThrProGlyLysProLysThrTh 1128
58531 ATACCCTGGAGCTG.....TATACGCTGGAGCAAAATCCACG... 58493
1128 rAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGln 1145
58492 .....A 58492
1145 exGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1161
58491 GCNAAGCTCCGCAACAGCAATGACAACTCGCGGACGACGAGTGGCTCCA 58442
1162 SerAsnProSer.....SerProGlyArgIleLysGlyArgLeuAsps 1176
58441 AAGAGACCGCGGCAACAGTCAATCATGATGATGAAATCCGCGAGTGAG 58392
1176 exSerGluMetAspHisSerGluAsnGluAsp..... 1186
58391 TGAGGTCCTTGGAAAATACTGAGAATGAATCTAACTAGTGTGTGAAATTT 58342
1187 .....TyrThrMetSe 1190
58341 GATTTCAGGAGAACAAATCTGTTGCAACATACGACGCGCTGACACTAGC 58292
1190 rSerProLeuProGlyLysLysSerAspLys 1200
58291 GAGCCCATGCCAAACAGACGCGCGGAGGAG 58261
```

seq_name: gb_pr5:AK026889

seq_documentation_block:

LOCUS AK026889 2079 bp mRNA PRI 29-SEP-2000

DEFINITION Homo sapiens CDNA: FLJ23236 fls, clone COL00725.

ACCESSION

AK026889.1 GI:10439854

VERSION oligo capping; fis (full insert sequence).

SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL00725.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

TITLE

NEDO human cDNA sequencing project

JOURNAL

Unpublished (2000)

REFERENCE

2 (bases 1 to 2079)

AUTHORS

Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE

Direct Submission

JOURNAL

Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES

source

1..2079

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="COL00725"

/clone_lib="COL"

/tissue_type="colon"

/note="cloning vector pME18SFL3"

304..681

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB1584.1"

/db_xref="GI:10439854"

/translation="MDLTKLVQEQPKGQSQRKRGHTASEDEQWPEKRLKEDI
LENEDEQNSPPKGRGRRPKPLGGGTPEKPTMTKTKSGSKSGPPAPEEEEEER
QSGNTEQSKSKQHRVSRRAQR"

BASE COUNT 759 a 354 c 339 g 627 t

ORIGIN

alignment_scores:

Quality: 925.50 Length: 214

Ratio: 5.057 Gaps: 1

Percent Similarity: 85.514 Percent Identity: 85.514

alignment_block:

US-09-512-581-2 x AK026889 ..

Align seg 1/1 to: AK026889 from: 1 to: 2079

1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluG1 1225

|||||

238 TCTGAATTGGAGAACGCTAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 287

|||||

1225 nGluGluLysLeuGlyMetAspLeuThrLysLeuValGlnGluGlnL 1242

|||||

288 GGAGGAGAAATTTAGGTATGATGACTTGACTAGTTGGTACAGGACACA 337

|||||

1242 ysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258

|||||

338 AACCTAAAGCGAGTCAGCGAAGTCGGAGGAGGAGGAGGAGGAGGAGGAG 387

|||||

1259 SerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLe 1275

60843 TCAGGCCCAACGACGCCAACATTCGCCAAGTCGGTCTCTATATAAGTTTTTC 60799

525IlePheSerLysValMetValIleThr.ArgAsnLeuP 537
::: ||| |||::: ::: ||| |||
60793 TTTTGATAGTTTGAACACAGTATGTTTCTTCTCAGACTGCTTC 60744

537 roAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu 553
||||||| ||||| ::::::::::: |||::: |||
60743 CAGATCCCTAAAGCGCAGAGTACCTTAACCCAGTTTGGCAACAACCTG 60694

554 GluAspAspGluLysIleArgLysGlnLeuGluValSerProTh 570
::: |||::: ::::::::::: |||::: |||
60693 CGAAAGATGCTCAGTCTCTACGATGCAATTAACATGTGCTTAAACGCGA 60644

570 rCySerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysL 587
||||||| |||::: ::::::::::: |||::: |||
60643 CGTAGCTGTGGGAGTGTGTGACACATGGGAGTCTCTCTGAAACACG 60594

587 euGlyAsnProLysGlnProThrAsnPropheLeuGluMetIleLysPhe 603
||||| ||| ::::::::::: |||::: |||
60593 TTGGCGCCACAGTCCAA...TCGAATCTGTATTACAACACAGTTAAGATG 60547

604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAl 620
||||||| |||::: |||::: |||::: |||::: |||
60546 CTGATGTAGCGGTGGCATCTAGTGTGTGACACAGGAGTCTATTGGCTG 60497

620 aleuIleLys..... 623
:|||||:::

60496 ACTTATTAGTAAATGTCTTTAAGTGGCCTTTTGCACAGCTTAATCCGG 60447

624GlnValAsnLysSerIleAspGlyThrAlaAsp 634
::: |||::: |||::: |||::: |||::: |||
60446 TAATCTTTTGTCTCAGCCTAATTGAACAATGCATCGACAAAGGTCCATG 60397

635 AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGl 651
||||||| |||::: ::::::::::: |||::: |||
60396 TGTGAAGAGATCGGAATTTTCGGCTCAGGAAGCGGCGAGCGGACTCAA 60347

651 uLeuLeu.Lys..... 654
:|||||:::

60346 GTTGTGGCCGTGAGTATAGAGATAATTTAAGTTATGATACAGAAATCCA 60297

655ValLeuSerPheThrHisProIleSerP 664
::: |||::: |||::: |||::: |||::: |||
60296 TTTGATGCTTTATACTTGCAGATGCTTTTATACGCTCTTTTCGGCTCAT 60247

664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
||||||| |||::: |||::: |||::: |||::: |||
60246 TCTTCCCGACTACTTCGGTGGCTATTAAATTTCCCTGCTCAGTTACGAG 60197

681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly.. 696
::: |||::: |||::: |||::: |||::: |||
60196 CAGGATTATGTTGCACCACTGGTGTCTAAAGACATACTCATCTGGGCGC 60147

697SerLysIleGluGluAspPheProHisIleArgSerAlaLeuL 711
||||||| |||::: |||::: |||::: |||
60146 CTATCAGCCACTATTGACGACCCCTACACCGGCTATTCTCAGCAGGCTGG 60097

711 euProVal.....LeuHisLysLysSerLysLysGlyProProArg 724
||||||| |||::: |||::: |||::: |||
60096 CTCGGTGTGCAAGGACTTCCACTT.....ATTGGAACCCCAAG 60057

725 GlnAlaLysTyrAlaIleHisCysIleHis..... 734
||||||| |||::: |||::: |||::: |||
60056 CAAGCAAGACGCGGTGCGATGTATTTTCGTAAACAGCCAGTCTGTCGCG 60007

735AlaIlePheSerLysLysGluThrG 743
||| |||::: |||::: |||
60006 TTCCACTGATGGACCAACAGTGGAGCTGGGAAGTCGCTCCACACTACAC 59957

743 In.....PheAlaGlnIlePheGluProLeuHisLysSer 754
|| ||| |||::: |||::: |||::: |||
59956 AAACAGTGCATCTATTTCACGAGATCATCTCGACGAGCTTCGCTCAAAA 59907

[illegible]

Align seg

Align seg

3394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe 410
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33617 GCGAGCGAGCTAGATAAGAAGTACAAAATTCGCAGGGATGCCGTAA 33666
tclyleuAlaGlnIleTyrlsLysTyrAlaLeuGlnSerAlaAlaGlyL 427
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33667 TGGCTGCGCTACATCTACAAGCGGCANNTTGGCAACCCAACCATCTAA 33716
427 ysAspAlaAlaLys...GlnIleAlaTrpIleLysAspLysLeuLeuHis 442
::: :::: ||| ::::: ||||| ||||| ||||| ||||| |||||
33717 GCACGTGGCCTTAAGTGCTAGCTAGGCTTGACTGGATTAAAGAACAGTACTACAT 33766
443 IleTyrrTyrrGlnAnsnSerIlcAspAspArgLeuLeuValGluArllePh 459
||||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
33767 GGATACTACAAGTGGGCTTGGAGGACCGCTGCTTGTGGAAACGCCCTACT 33816
459 eAlaGlnTyrrMetValProHisAsnLeuGluThrThrGlnArgMetLysC 476
::: : ||||| ||||| ||||| ||||| ||||| ||||| |||||
33817 TATCACCTGCGCTGGTTCCCTATAACTAGCTCCGGAAGAGCGCATGAAGA 33866
476 ysLeuTyrrTyrrLeuTyrrAlaThrLeuAspLeuAsnAlaValLysAlaLeu 492
||||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
33867 AGCTATATCATTTGTCAGCGCATTCGATGCGCAATAGCGTCACTGATGTG 33916
493 AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe 509
||||| ::::: ||| ||| ||||| ||||| ||||| ||||| |||||
33917 CTCGAGCTGCAGAGAAGCAAAATGAAGACACAGCAATAGCGTCACTGATG 33966
509 uLeuAspLeuIleLysGln.....ProlSynThrAspAlaSerV 522
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
33967 GATCAAGCTACACCACCTTAGGAGTTCACCGCGCGGTACTTAGTCAGC 34016
522 alLysAla..... 524
::: |||||
34017 TCAGCGCCAAGCAGCGCCAACATTCGCAAGTGGTCTATATAATAGTTTTT 34066
525IlePheSerLysValMetValIleThr.ArgasnLeuP 537
::: ||| ||||| :::: ||||| ||||| ||||| |||||
34067 TTTTGATAGTTTGAACACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 34116
537 rOAspPrGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu 553
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34117 CAGATCCCCTAAGCGCGCAGTACCTAACCCAGTTAGCAACAACCTG 34166
554 GluAspAspGluLysIleAryLysGlnLeuGluValLeuValSerProTh 570
::: ||| ||||| :::: ||||| ||||| ||||| ||||| |||||
34167 CGAAAAGATGCTCAGCTCCTACAGTGAATTAACATGTGCTCTTAACGCGA 34216
570 rCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysL 587
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34217 CGTAAGCTGTGGAGTGTGCTGACAGTGGAGTCTCTCTGAAAAAGC 34266
587 euGlyAsnProLysGlnProThrAsnPropHeLeuGluMetIleLysPhe 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34267 TTGGCGCCACGTCCAA...TCGAATCTGTATTACACACAGTTAAGATG 34313
604 LeuLeuGluArylAlaProValHisIleAspThrGluSerIleSerAl 620
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34314 CTGATTGACGCGTGGCATCATGTGTTGACAGAGGAGTCTATTGGCGT 34363
620 aleutileLys..... 623
||||| |||||
34364 ACTTATTAGTAAATGCTTTTAAGTGCCTTTTGGCCAGACTTAATCCGG 34413
624GlnValAsnLysSerIleAspGlyThrAlaAsp 634
::: ||||| |||||
34414 TAATCTTTTGTCTAGCCTAATTGAACAATGCATCGAGAAAGGCTCCATG 34463
635 AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGl 651
||||| ::::: ||||| ||||| ||||| ||||| ||||| |||||
34464 TGTGAAGATGCGAATTTCCGCTCAGAACGCGCGACGCGCATCAA 34513

551 uLeuLeu...Lys..... 654
34514 GTTGTGGCCGTGAGTAGAGATAAATTAAAGTTATATGATACAGAAATCCA 34563
655ValLeuSerPheThrHisProIleSerP 664
34564 TTTGATTGCTTTATACTTGCAGATGCTTTCATAGCTCTTTTCGGCTCAAT 34613
664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
34614 TCTTCACCGATACTTCGCTGCGTCATTTAAATTTCCTGCTCAGTTACGAG 34663
681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly... 696
34664 CAGGATATGTTGCACCACTCGTCTAAAGACACTAATCTATCTGGGCGC 34713
697SerLysIleGluAspPheProHisIleArgSerAlaLeuL 711
34714 CTATCAGCCACTCAATTCAGACCCCTACACGGCTATTCTCGACGAGTGG 34763
711 euProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLys 727
34764 CTCGCGTGTCAAGACTTCGCATTTATGGAAACCCCAAGCAAGCAAG 34813
728 TyrAlaIleHisCysIleHis..... 734
34814 CACGGGTGCCATGATATTTTCGTAAACAGCCAGCTCGTGGCTTCCACTGA 34863
735AlaIlePheSerSerLysGluThrGln..... 743
34864 TCGAGCAACAAGTGGAGTGAAGTGCCTCAACAACACTACAACAACAGTGC 34913
744PheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
34914 ATCCTATTTTCAACGAGATCATCGAGCGCTTCGCTAAACTGACTGACACA 34963
758 SerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAl 774
34964 ..AACTGTGAGCATCAGCGCACAAAGATTGTGACCTTGGGTGCATTTGC 35010
774 aLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaI 791
35011 CTACAAATATGCCACAGGCTTCTTCAACGCCCTTAAGAANAATATGATTTGC 35060
791 hrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys 807
35061 GACGAATGTCAAAGAGCTGCTTATCCAAGAA...GTTCTCTGCACGCGA 35107
808 ThrThrLysLeu.....TrpValProAspGluGluValse 819
35108 GACTACGAACGTGCCAGAGGACAGTCATTTGGTGTGCCAAGAGAATAC 35157
819 rProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL 836
35158 GCCGACACTCTATTAAGTGTGATGCGCTCAAGGCTATGCCAGGTGGC 35207
836 euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
35208 TTTTGGGATTCCGTACCCGAT...GAGCACGCTGCCCAAGAAACATTCGA 35254
853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLy 869
35255 ATGTGGCGGGCTTTTAAACACAGGAGCGCATTTGTCGGCCACAGACCG 35304
869 sIleSerLysProAspMetSerArgLeuArgLeuAlaGlySerAlaI 886
35305 TCTTTGCGGTGCCAGAAATCTTGCTCGCTCGCGGCGACCTTCGCGCCA 35354
886 leValLysLeuAlaGlnLupProCysTyrHisGluIleIleThrLeuGlu 902
35355 TGCTCAAGTGTGGCAACAAAGGCGGTAGGTGATCAGTACAGCGCGTGA 35404
903 GlnTyr...GlnLeuCysAlaLeuAlaIleAsnAspGlu..... 914

Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, J., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006

2 (bases 1 to 262731)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7303489.

FEATURES

Location/Qualifiers

1..262731
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 complement(join(5117..5856,6026..6362,6424..7856,
 8828..9717,17398..17559,17629..17763,22517..23464))
 /gene="otk"
 /note="Nucleotide sequence of the Celera sequence differs
 from the published sequence for this transcript."
 /product="CT25769"
 /db_xref="FLYBASE:FBan0008967"
 /db_xref="FLYBASE:FBgn0004839"
 complement(<5117..>23464)
 /gene="otk"
 /note="otk"
 /map="48D8-48E2"
 /note="CG8967"
 /db_xref="FLYBASE:FBan0008967"
 /db_xref="FLYBASE:FBgn0004839"
 complement(join(5764..5856,6026..6362,6424..7856,
 8828..9717,17398..17559,17629..17763,22517..22568))
 /gene="otk"
 /note="otk gene product: Nucleotide sequence of the Celera
 sequence differs from the published sequence for this
 transcript"

/codon_start=1
 /db_xref="FLYBASE:FBan0008967"
 /db_xref="FLYBASE:FBgn0004839"
 /protein_id="AAF58596.1"
 /db_xref="GI:7303541"
 /translation="WTARMISICGLVWALMWASVLASSRRFORVPQSQSVVENESVKF
 ECESDTSIEUHWYDNGHRIADYKRVHQIGSNLHIEAVRTEDEGVNVCYATNLAS
 GARBAAPKLSVLYLESASVOLGSSNRNELLKCHVEGSGDLEPLEIEWRNSKLL
 STWRNVDQHRLLIROPGSEDGLYRCTASNAAGYVSGKYQSKCLPLRPLRR
 NKNKMSWDKOTFLCRKGGAGLEALPAAPEDLRIVQGPIGOSIIEKEPTALTC
 LYEPLDELKRIOLRWKDKGLROVELGGSAPICGHSFDGSKDALLREDARLVLHK
 QNGTLSFIITASDAGYQCOQLEAHNPINSSPGILEVIEQLKFVQPTSKNLELDA
 VVAKVHCQAQGTPTQVQWRDGGENTTLPDHVEVDANGTLIFRNWNSHRNGNYTCLAT
 NSQOINATVAINTVVPVPPVPIETSEQGTVMHCOAIGDPKPTIOWDRLKY
 LSENNTRERFRLENTLEIRNVQVEDESGYCTIGNAGLKREDVQLVVKTTGDFG
 APESGGDGLVLTFTVVALAYIVLVGLMWCYRQARKARNDLSTKEAGG
 DQPDVAGKSGSEPCLSKOHNGHSKRSKSGDAOKSDDTACSQOSRASKSAHIY
 EQALPRSGSELTOIGRGEFGDVFKLKATLVTSISDKDADTEKHNSNENSGSGS
 GSGSTTSLNEKRKRSTKMDIDIEIEEEDQDQHNQSGLEQLVLVKALNKVKDQAOQ

mRNA

EFRRQLDLRAISHKGVVFLGCLREKDPHYMVLEYTDWGLKQFLATACKVNTATA
 GSSPPPLTTSSQVLAVAYQIARFARFTRDLATRCNVISSEEIVKVSYPALC
 KDKYSRVYKHKRNTLLPIRWLAPGICQIEDVYTKSDIFAYGVVWELFNPQATKLPHEE
 LTNQVVQVQSQAGSLENSVAREATPDSLREILLSCWNSNPKRPSFSQLGAALSRAMQS
 AEK"
 complement(join(<27067..27801,27890..28383,28455..>28527))
 /gene="CG8964"
 /product="CT25760"
 /db_xref="FLYBASE:FBan0008964"
 /db_xref="FLYBASE:FBgn00033674"
 /evidence=not_experimental
 complement(<27067..>28527)
 /gene="CG8964"
 /map="48E1-48E2"
 /db_xref="FLYBASE:FBan0008964"
 /db_xref="FLYBASE:FBgn00033674"
 /evidence=not_experimental
 complement(join(27067..27801,27890..28383,28455..28527))
 /gene="CG8964"
 /note="CG8964 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008964"
 /db_xref="FLYBASE:FBgn00033674"
 /evidence=not_experimental
 /protein_id="AAF58595.1"
 /db_xref="GI:7303540"
 /translation="MGLNGRRRLGCGLFFCLIFCSYAAPAFEDVISISRGPKSIT
 IKEOSSLQPCDYOLPDGYLOKSSVILRWKDKSTLROVELGRMDSTTSEPOLETMLR
 EDSVLSKDSGALQFTSVLASDAGYQCOLVIDDSVAASSGVLLVLEQLKVPQP
 TSKNLELGTLSKVHCKAGQTPAQVKWRETQLPLPNVTQNGTLIFNQVSNQRGQ
 YTCIASNSQGITATVSNVVAAPKFSVPPPEGIEVAAGTAVIHCOAIGEPKPTIQW
 DKULTYLNENNTDPERFSLMENGTELRNVRPEDEGRYGTIGSAGLKREDVLLVLK
 SKSASIVTRIIVIIICLAFVLVLGLKVMVRYRRLKVKQLGLEDGHVNGPTDQ
 EHDHSENPCLTSSSNKLSKLESTILBOESQVADDIV"
 complement(join(29253..29670,29731..30211,30362..30624,
 30681..30790,30862..31313))
 /gene="CG8889"
 /product="CT25522"
 /db_xref="FLYBASE:FBan0008889"
 /db_xref="FLYBASE:FBgn00033675"
 complement(<29253..>31313)
 /gene="CG8889"
 /map="48E1-48E2"
 /db_xref="FLYBASE:FBan0008889"
 /db_xref="FLYBASE:FBgn00033675"
 complement(join(29483..29670,29731..30211,30362..30624,
 30681..30720))
 /gene="CG8889"
 /note="CG8889 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008889"
 /db_xref="FLYBASE:FBgn00033675"
 /protein_id="AAF58594.1"
 /db_xref="GI:7303539"
 /translation="MAQSSWQPIDCKLDCNTRLLIADPQILGNSYDRSSPSLARYD
 SDRYLAKTFERALFTQPHITVFLGDLDEGNIAIYQYQVRRRIYQNKYKRR
 VHVPDNDIGGENDYISNSQRRFENESEDLFDNDRNRFKINMLDLDSNPDR
 DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIIFSGHWHESRIIFYPSTKVINYF
 ENSVRHFDLKALKEQHSYLEIMVPTCSYRMKSKIGIGYAVLENYLSYTVLQPNR
 FILLFTYFWGLFVVCGVFVKMMTRCPFRVAKRQTLYNRYSTIPQF"
 join(31631..31871,32192..32294,32356..32741,32805..33012,
 33071..33239,33303..34044,34109..34372,34430..34523,
 34586..35431,35497..35633,35695..35774,35836..36101,
 36174..36464,36528..>36620)
 /gene="CG17509"
 /product="CT32429"
 /db_xref="FLYBASE:FBan0017509"
 /db_xref="FLYBASE:FBgn00033676"
 /evidence=not_experimental
 /protein_id="AAF58594.1"
 /db_xref="GI:7303539"
 /translation="MAQSSWQPIDCKLDCNTRLLIADPQILGNSYDRSSPSLARYD
 SDRYLAKTFERALFTQPHITVFLGDLDEGNIAIYQYQVRRRIYQNKYKRR
 VHVPDNDIGGENDYISNSQRRFENESEDLFDNDRNRFKINMLDLDSNPDR
 DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIIFSGHWHESRIIFYPSTKVINYF
 ENSVRHFDLKALKEQHSYLEIMVPTCSYRMKSKIGIGYAVLENYLSYTVLQPNR
 FILLFTYFWGLFVVCGVFVKMMTRCPFRVAKRQTLYNRYSTIPQF"
 join(31631..31871,32192..32294,32356..32741,32805..33012,
 33071..33239,33303..34044,34109..34372,34430..34523,
 34586..35431,35497..35633,35695..35774,35836..36101,
 36174..36464,36528..>36620)
 /gene="CG17509"
 /product="CT32429"
 /db_xref="FLYBASE:FBan0017509"
 /db_xref="FLYBASE:FBgn00033676"
 /evidence=not_experimental
 /protein_id="AAF58594.1"
 /db_xref="GI:7303539"
 /translation="MAQSSWQPIDCKLDCNTRLLIADPQILGNSYDRSSPSLARYD
 SDRYLAKTFERALFTQPHITVFLGDLDEGNIAIYQYQVRRRIYQNKYKRR
 VHVPDNDIGGENDYISNSQRRFENESEDLFDNDRNRFKINMLDLDSNPDR
 DNNADRLRIGVSHAPLLIGGPLLRAIISLDLPHIIFSGHWHESRIIFYPSTKVINYF
 ENSVRHFDLKALKEQHSYLEIMVPTCSYRMKSKIGIGYAVLENYLSYTVLQPNR
 FILLFTYFWGLFVVCGVFVKMMTRCPFRVAKRQTLYNRYSTIPQF"

gene

CDS

mRNA

gene

CDS

mRNA

gene


```

147549 AGCTCTACACGATATCACCGAAATTCGCGCTTCGAAATCACGACTT 147598
      ::::: ||:::||||| ||:::||||| ||:::||||| ||:::|||||
362 oGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaL 379
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
147599 TGACGAAGTGGTCCGCCACGAGGTGGTAATGCTATTGTGGAACAGTCCA 147648
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
379 yLysAspIleLeuLeuValAsnAsp.....HisLeuLeuAsnPheVal 393
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147649 AGCGGACTTCACCCCTCGTTCTCGAAGCGCCGCTACTAGAAATTGTG 147698
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe 410
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147699 CGGAGCGGACGCTAGATAAGAGTACAAATTCGCGGGATGCGATGAA 147748
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
410 tGluLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL 427
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147749 TGGTCTGGCTACATCTACAGCGCGCAATTCGGAACCCACGATCTAA 147798
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
427 yAspAlaAlaLys...GlnIleAlaTrpIleLysAspLysLeuLeuHis 442
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
147799 GCACGTGGCCTTAAGGTCAGGTTGACTGGATTAAGAACAGATACTACAT 147848
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
443 IleTyrTyrGlnAsnSerIleAspArgLeuLeuValGluArgIlePh 459
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147849 GGATACTACAAAGTGGCTTGGAGGACCGCTCGTGTGGAACGCGCTACT 147898
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
459 eAlaGlnTyrMetValProHisAsnLeuLeuGluThrThrGluArgMetLysC 476
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147899 TATCACCTCCCTGGTTCCTATATACTAGCTCCGGAGAGCGCATGAAGA 147948
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
476 yLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeu 492
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147949 AGCTATATCATTTGTAGCGATCTCGATGCCAATGCCACCAAGGCTTTC 147998
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
493 AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe 509
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
147999 GTGAGCTCGAAGAAGCAACAAATGAAGACACGCAATACCGTCAGTATTG 148048
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
509 uLeuAspLeuIleLysGln.....ProLysThrAspAlaSerV 522
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
148049 GATCAAGCTACACCACTCTAAGGAGTTTCACGCGCGCTTACTAGTCAC 148098
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
522 aLysAla..... 524
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
148099 TCAGGCCCAAGCAGGCGCAACATTGCCAAGTGGTCTATAATAAGTTTTC 148148
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
525 .....IlePheSerLysValMetValIleThr.ArgAsnLeup 537
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148149 TTTTGATGTTGTACCAAGTATGTTGTTTTTTTCTTCAGACTGCTTC 148198
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
537 roAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeu 553
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148199 CAGATCCCTTAAAGCGGAGAGTACTTAACCCAGTTTAGCAACAACCTG 148248
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
554 GluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProTh 570
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
148249 CGAAAGATGCTCAGCTCTACAGTGCATTAACATTTGCTTTAAACGCGA 148298
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
570 rCysSerLysLysGlnAlaGluLysValArgGluIleThrLysLysL 587
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148299 CGTAAGCTGCGGAGTGTCTGACACGATGGGAGTTCTCTGAAAAAGC 148348
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
587 euGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPhe 603
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148349 TTGGCGCCACCTCCAA...TCGAATCTGATTACAACACAGTAAAGATG 148395
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAl 620
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148396 CTGATTGACGCGTGGCATCAGTGATGGTGGACAAGGAGTCTATTGGCGT 148445
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
620 aLeuIleLys..... 623
      :::::||||| ||:::||||| ||:::||||| ||:::|||||

```

```

148446 ACTTATTAGGTAAATTGCTTTTAAGTGCCTTTTGGCCAGACTTAATCCGG 148495
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
624 .....GlnValAsnLysSerIleAspGlyThrAlaAsp 634
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
148496 TAATCTTTTGTCTACGCCCTAATTGAACAATGCATCGAAGAGGTCCATG 148545
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
635 AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGl 651
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148546 TGTGAAGAGATCGGAATTCGGCTCAGGAAGCGGCGAGCGGACTCAA 148595
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
651 uLeuLeu...Lys..... 654
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148596 GTTGTGGCCGTGAGTATAGAGATAATTTAAGTTATGATACAGAAATCCA 148645
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
655 .....ValLeuSerPheThrHisProIleSerP 664
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
148646 TTTGATTGCTTTATATCTTCAGATGCTTTTCATACGTCCTTTTCGGCTCAT 148695
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148696 TCTTCACCGATACTTCGCTGCGTCATTTAATTTCCCTGCTCAGTTACGAG 148745
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly... 696
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148746 CAGGATTATGTCACCACTGGTGCTAAAGACACTAACTCATCTGGGGCG 148795
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
697 .....SerLysIleGluGluAspPheProHisIleArgSerAlaLeuL 711
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148796 CTATCAGCCACTCATTCAGACCCCTACACCGGCTATTCTCGACGAGCTGG 148845
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
711 euProValLeuHisLysSerLysLysGlyProProArgGlnAlaLys 727
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148846 CTCCTGCTGTCAAGGACTTCGCACTATTGGAACCCCAAGCAAGCAAG 148895
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
728 TyrAlaIleHisCysIleHis..... 734
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148896 CACGCGTCCGATGATTATTTCTAAACACCCAGTCGTCGGCTTCCACTGA 148945
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
735 .....AlaIlePheSerSerLysGluThrGln..... 743
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148946 TGGAGCAACAAGTGGAGCTGGAAGTGCCTCCCACTACACAACACAGTGC 148995
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
744 .....PheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
148996 ATCCCTATTTTCAACGAGATCATCGAGACGCTTCGCCATAAACTGACACCA 149045
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
758 SerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAl 774
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
149046 ...ACTGTGACATCAGCGCACAAAGATTGTGACCTTGGTCAACATTGC 149092
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
774 aLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlar 791
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
149093 CTACAATATGCCACAGGCTTCTTAACGCCCATTAAGAATATGATTGCGC 149142
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
791 hrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys 807
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
149143 GACGAATTGTCAAAGAGCTGTATTCCCAAGAA...GTTCCTGCGCAGCGA 149189
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
808 ThrThrLysLeu.....TrpValProAspGluGluValSe 819
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
149190 GACTAGCACTGCCAGACAGACTGATGCTGTGCCCCAAGAGAAACTACC 149239
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
819 rProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL 836
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
149240 GCCGACACTCTATGTAAGCTGGATGCGCTCAAGGCTATGGCCAGGTGGC 149289
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
836 euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
      ||:::||||| ||:::||||| ||:::||||| ||:::|||||
149290 TTTTGGGATTGCGTACCGGAT...GAGCAGCTGCCCAAGAAAACTCCGA 149336
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLy 869
      :::::||||| ||:::||||| ||:::||||| ||:::|||||
149337 ATGCTGGCGCGCTTGTGTTAAACCAACGAGCGGATTGCTCGCCAGAACCG 149386
      :::::||||| ||:::||||| ||:::||||| ||:::|||||

```



```
869 sileserLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaI 886
25110 WTTCGCGTGCAGAAATCTGGCTCGCGGCGAGCTGCGCCA 25159
886 leValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGlu 902
25160 TGCTCAAGTGTGCAACAAAAGGCGGTAGTGATCAGTACAGCGCTGAG 25209
903 GlnTyr...GlnLysCysAlaLeuAlaIleAsnAspGlu..... 914
25210 CAGTATTTCAGCTTCCCGAGCTGATGGTGAGTAGAGACATCATCTTGT 25259
914 ..... 914
25260 GCTTAGAAGCTTAGTCTCAATGCTATCTCGGTGTTCCCGTAGGCTGATCC 25309
915 ..CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeu 930
25310 GGTGCCA.GAAGTTCGGGAATCTTTCGTCGCAAGCTGCACAAAGGATTA 25358
931 SerArgLeuArg.....LeuProLeuGluTyrMetAlaIleCys 943
25359 AGCAGAAGTTTCCCGAGGAAGTGTTCGCGCTGGAGCTTCATGGCTGTA 25408
943 salaleuCysAla.....LysAspProValLysGluArgA 955
25409 TGTGTCGTGCTGTAGAGACTGAGGAGAGTGTAGTTCAGAAACCA 25458
955 rgAlaHisAlaArg.GlnCysLeuValLys..... 964
25459 TCCGATCCGTTGGAGAGTATTTGATAAATATTTTCAACAGATTGCAAGA 25508
965 .....AsnIleAsnValArgArgGluTyrL 973
25509 CCTTGTGCTCACTATGCAGAACGATGATAAACAAACGGCGGAATATC 25558
973 euLys..... 974
25559 TCAAGACTGTGCTATGACATGATAGTGTGTTTACTTTTGGAAATATATTTG 25608
975 .....GlnHis.AlaAlaValS 980
25609 TGTATTTTAACCATTAATCAATATTATCCAGCTCCCGACAGCTCAACGG 25658
980 erGluLysLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
25659 AGTCACAATCATACACATACTACCTGACTACATGCTGCTTCGCTATT 25708
997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluI 1013
25709 CCGGTGCTGGTCCAGATCCAGCTTCAGAAATCAGAGGACTGATACA 25758
1013 nLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMeta 1030
25759 GCTGCGCAGATGGAGAGTGGCTGGCTTCATTCTGGAGCGCTGATGG 25808
1030 laLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046
25809 CCAAAAGCAGAAAAGTTGTTCATAGCTTCTACAAGCAGCTGCTGCAAGCTG 25858
1047 IleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn... 1062
25859 ATAAAGCATCGCGAGTTTCAGTCTGGGG...TCGGACAAGCGCGCAACATTA 25905
1062 ..... 1062
25906 TGTATCTCACTCAATCATTTTCCAGGCTTTCCCTTCTAAATATATTTA 25955
1063 .....GluLysLeuTyrThrValCysAspValAla 1072
25956 TACATATATTTTCGTTATCAGAAAATGTGGCGCTGTCGATCTTGGC 26005
1073 MetAsnIleIleMetSerLys.....SerThrTh 1082
```

```
26006 ATGTACATATATCGACTCCAAAGTTTCAGTCCATTTGATGCAACACGAGCAC 26055
1082 rTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheT 1099
26056 CTTTTCATGCGCGTG.....GCTTTCAGGAAATGTTATATA 26093
1099 hrGlnProAsp...LysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
26094 AAGAGCCTGCCGTTGCGAATTTCCAAAACAATGACGCTCTATATACCGCTG 26143
1115 GluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValle 1131
26144 GACGTG.....TATACGTGGGAGCAAAATCCACG..... 26173
1131 uGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrL 1148
26174 .....ACCAAAGCTG 26183
1148 ysSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPro 1164
26184 CGCAACAGCAATGACAACGTCGCGAGCAGCAGTGGCTCCAAAGAGACGG 26233
1165 Ser.....Ser.ProGlyArgIleLysGlyArgLeuAspSerSerGluM 1179
26234 CGGGAACAGCTCAATCATGATGATGAAATCCGACGAGTGAGTGCTCT 26283
1179 etAspHisSerGluAsnGluAsp..... 1186
26284 TGGAAATACTGAGATGNACTACTAGTGTGTAATTTTTCATTTCAGG 26333
1187 .....TyrThrMetSerSerProLe 1193
26334 AGAACAACTGTTCGACACATACGAGCGCTGACACTACGAGGCCCCATG 26383
1193 uProGlyLysLysSerAspLys 1200
26384 GCCAACGAACCGCGCAGGAG 26405
seq_name: gb_inl.AC007474
seq_documentation_block:
LOCUS AC007474 192763 bp DNA INV 28-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
AC007474, complete sequence.
ACCESSION AC007474
VERSION AC007474.5 GI:13162477
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 192763)
Celisner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ikebawam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirska,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 49A-49B
Unpublished
2 (bases 1 to 192763)
Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
```


23272 AGTCTTCAGCAGCATATACACGAGAAATTCGGCTTCGAAATCACGATCT 23321
 362 oGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaL 379
 23322 TGACGAAGTGGTCGCGCAGCAGGTTGGAATGGCTATTGTGGAACACTGCCA 23371
 379 yLysAspIleLeuLeuValAsnAsp.....HisLeuLeuAsnPheVal 393
 23372 AGCGGACTTCACCCCTCGTCTCGAACGCCCGCATCTACTAGAAATGTG 23421
 394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMe 410
 23422 CGCGAGCGGCGGTAGATAAGAAAGTCAAAAATTCGAGGGATCGATCAA 23471
 410 tGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL 427
 23472 TGGTCTGGCCTACATCTACAGGCGCGCAATTCGGAACCCCAAGCATCAA 23521
 427 yAspAlaAlaLys...GlnIleAlaTrpIleLysAspLysLeuLeuHis 442
 23522 GCACCTGGCCTTAAGTTCAGGTTGACTGGATTAAACAACAGATACAT 23571
 443 IleTyrTyrGlnAsnSerIleAspArgLeuLeuValGluArgIlePh 459
 23572 GGATACCTACAAAGTGGCTTGGAGACCGCTGCTTGTGGAACGCTACT 23621
 459 eAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysC 476
 23622 TATCACCTGGCTGGTTCCTTAACTAGTTCGGAAGCGCATGAAGA 23671
 476 yLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeu 492
 23672 AGCTATATCATTTGCTAGGCGATCTCGATGCCAATGCCACCAAGCCTTC 23721
 493 AsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLe 509
 23722 GTCAGCTGCAGAAACAAATGAACACACGCAATACCGTTCAGTGATTG 23771
 509 uLeuAspLeuIleLysGln.....ProLysThrAspAlaSerV 522
 23772 GATCAAGCTACACCACTCTAAGGAGTTTCAGCGCGCGTACTTAGTCAGC 23821
 522 alLysAla.....IlePheSerLysValMetValIleThr..ArgAsnLeuP 524
 23822 TCACGCCCAAGCAGGCCAACATTCCTCAAGTGGTCTATAATAAGTTTC 23871
 525IlePheSerLysValMetValIleThr..ArgAsnLeuP 537
 23872 TTTTGATAGTTTGTAAACAGTGTGTTTTTTTTTTCTTCAGACTGCTTC 23921
 537 roAspProGlyLysAlaGluAspPheMetLysLysPheThrGlnValLeu 553
 23922 CAGATCCCTTAAAGCGGAGTACCTAACCCAGTTTAGCAACAACCTG 23971
 554 GluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProTh 570
 23972 CGAAAAGATGCTCAGCTCCACGATGATTAACATTGCTCTTAAACCGCA 24021
 570 rCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysL 587
 24022 CGTAAGCTGTGGAGTGTGTGCACACGATGGGAGTTCTCTGAAAAAGC 24071
 587 euGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPhe 603
 24072 TTGGCGCCCGCTGCTCAA...TCGAATCTGTATTACAAACACAGATTAAAGT 24118
 604 LeuLeuGluArgIleAlaProValHisIleAspPheThrGluSerIleSerAl 620
 24119 CTGATTGAGCGCGTGGCATCAGTGATGGTGACACAGGAGTCTATTGGCGT 24168
 620 aLeuIleLys..... 623
 24169 ACTTATTAGTAAATTCGTTTAAAGTGCCTTTTTCGACAGACTTAATCCGG 24218

624GlnValAsnLysSerIleAspGlyThrAlaAsp 634
 24219 TAATCTTTTTCCTCAGCCTAATTGAACAATGCATCGAAGAGGTCTCATG 24268
 635 AspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuG1 651
 24269 TGTGAAGAGATCGAAATTCGGCTCAGGAAGCGGCGGCGGACTCAA 24318
 651 uLeuLeu...Lys..... 654
 24319 GTTGTTCGGCTGAGTAGAGATAATTTAAGTTATGATACAGAAATCCA 24368
 655ValLeuSerPheThrHisProIleSerP 664
 24369 TTTGATTGCTTTATACCTGGCAGATGCTTTTCATAGCTCTTTTCGGCTCAT 24418
 664 heHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
 24419 TCITCACCGATACCTTCGCTGCGTCATTTAATTCCTGCTCAGTTACGAG 24468
 681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly... 696
 24469 CAGGATTATGTCACCACTGCTGCTAAAGACACTAACTCATCTGGGCG 24518
 697SerLysIleGluGluAspPheProHisIleArgSerAlaLeuL 711
 24519 CTATCACCCACTCATTCAGCACCTTCACCGCTATTCTCGACGAGCTGG 24568
 711 euProValLeuHisHisLysSerLysGlyProProArgGlnAlaLys 727
 24569 CTCGGGTGCAAGGACTTCGCCTTATTGGAACCCCAAGCAAGCAAG 24618
 728 TyrAlaIleHisCysIleHis..... 734
 24619 CACGCGTGGCTGATTTTCGTAACACGACGCTCGTGGCTTCCACTGA 24668
 735AlaIlePheSerLysLysGluThrGln..... 743
 24669 TGGAGCAACAAGTGGAGCTGGAGTGGCTCCACAACTACACAACAGTGC 24718
 744PheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
 24719 ATCTATTTTCAACGAGATCATCGAGAGCTTCGCTTAAACTGACACCA 24768
 758 SerAsnLeuGluHisIleThrProLeuValThrIleGlyHisIleAl 774
 24769 ...AACTGTGAGCATCAGCGCACAAAGATTGTGACCTTGGGTCAATTGC 24815
 774 aLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAla 791
 24816 CTACATATCCACAGGCTTCTTAACGCTTAAAGAAATATGATGCTGC 24865
 791 hrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys 807
 24866 GACGAATTGTCAAAGAGCTGCTTATCCAAGAA...GTTCTGCGCAGCGA 24912
 808 ThrThrLysLeu.....TrpValProAspGluGluValSe 819
 24913 GACTACGAAGTGCACGAGCAGTGTGTTGCTGCCAAGAGAAACTACC 24962
 819 rProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL 836
 24963 CGCGGACACTCTATGTAGCTGGATGCGCTCAAGGCTATGCGCAGGTGGC 25012
 836 euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
 25013 TTTTGGGATTGCGTACCAT...GAGCACGCTGCCCAAGAAACATTCCGA 25059
 853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLy 869
 25060 ATGCTGCGCGCTTGTGTTAAACCAACGAGGCGATTGTCGCGCCAGAACCG 25109

2190 GTAAAGATAAGGAA.....ATTAATTCTGATCAGGCTACC..... 2225

1238 lGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisT 1255

2226CAGGGCACATCATCAGCATGTGACCGGAGGAAGAAGACAGTAA 2268

1255 hrAlasSerGluSerAspGlu...GlnGlnTrrProGluGluLysArgLeu 1270

2269 CAGCAGCTGGTGCAGAGAATAATCCAAACAAAAACAGATGAGAAA..... 2312

1271 LysGluAspIleLeuGluAsnGluAspGluGlnAsnSerPro...ProLys 1286

2313GTAGTGAATCGGGAGCCTCCGCCCCCTC 2341

1286 sLysGlyLysArgGlyArgProProLysProLeuGly...GlyGlyThrP 1302

2342 CAACCCAGGAGGAGCGTCGACCCCACTCTGAATCTCAGGCGCAATGCTA 2391

1302 roLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLys 1318

2392 CCAAAAATGATCATCTAAATAAACCTATTAAACAGGGAAGGAGAGAGCT 2441

1319 ...SerGlyProProAlaPro.....Glu 1325

2442 GCAGTGGTGCAGGAGAGCCCTGGGGTGTGGGAAGCAGGTAATGCAAAAGC 2491

1326 GluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys..... 1340

2492 ACCCAAACATGCAAGATTATGCCCCAAAAGGCGACACCCAGCAAGAGACAA 2541

1341SerLysSerLysGlnHis.....ArgValSerArgArgAlaG 1353

2542 TTGACTTACAAAGGTAAAANTGCATTTCCAAAGGGAGAGAAATGAAGGCCA 2591

1353 lncInArgAlaGluSerProGluSerSer 1362

2592 AACAGAAG...CAGGCTCCAGCTTCTGCA 2617

seq documentation block:

DEFINITION *Drosophila m*

ACCESSION AC007475
VERSION AC007475.7

| | |
|--------|------------|
| SOURCE | fruit fly. |
| ORIGIN | fruit fly. |

Pterygota; N

REFERENCE
AUTHORS
I (bases I
Celniker, S. F.

Rogers, Y., A

Douson, K.: /
Ferreira, S.: /

Ibegwam, C.,
McIntosh, E. C.

Phouanenavon

TTTLE
Zaveri, J. S.,
Sequencing c

REFERENCE

Doyle, C.M.,

Hoskins, R. A.


```

gene      1. .2558
/ gene="KIAA0648"
CDS       <1. .2558
/ gene="KIAA0648"
/ codon_start=3
/ product="KIAA0648 protein"
/ protein_id="BA031623.1"
/ db_xref="GI:3327110"
/ translation="LYLYLASDPNAVNALNEMWKKMLRSHVRELLDLHKQPTSEA
NCSAMFGKLTIAKLNLPDPGKADDFVKFNQVLGDDKLRSLQLELLISPTSCCKQADI
CVREIARKLANPOPTNPFLEMYKFLLEIAPVHIDSEAIASLVKLMKNSIEGTADDE
EEGVSDPTAIRSGLELLKLVLSFTHTSFHSAETYESLQCLRMEDDKVAEAAIQIPEN
TGHKIEDLQIIRSTLPIILHQRKRGTPHOAKQAVHCIIHAIFTNKEVQLAQIPEPLS
RSLNADVEQLITPLVSLGHISMLAPDQFASPMKSVVAFIVKDLMLNDRSTGEKNK
LWSPDEVSPEVLAKQAIKLVLRLMGKMNQSKSANSTLRLSAMLVSEGDITEOK
RISKSDMLRLAAGSAIMKLAQPCVHEIITPEQFOLCALVINDECYQVRQIPAOKL
HKALVKLLPLEYMAIFALCAKDPVKERRAHARQCLKNISIRREYIKQNPMAITEKLL
SLLEPVVPMYHLLADHPDFTRSQVDQDLRIKELFMFLEVLMTKNENNSHAFMKK
MAENIKLTRAQSPDESKTNEKLYTCDVALCVINSKSALCNADSPKDPVLPKFFFTQ
PEKDFCNKSYIISSEETRVLLTGKPGAGVANKPLSATGRKPYVRSTGTETGSI
NVNSELNPTGNRSRQSEAAETGVSENEENPVRIITSVTPVKNDIPVKNKEINSQQA
TQGNISDRGKRTVTAAAGAEIOQKTDEKVDSEGPAPSKPRGRPRKRSSESQGNATK
NDDLKPIKGRRAAAGQESPGGLEAGNAKPKQLDLAKAAPAERQIDLQR"
BASE COUNT 1690 a 931 c 1037 g 1519 t
ORIGIN

alignment_scores:
Quality: 2704.50 Length: 910
Ratio: 3.675 Gaps: 15
Percent Similarity: 80.879 Percent Identity: 60.549

alignment_block:
US-09-512-581-2 x AB014548 ..
Align seg 1/1 to: AB014548 from: 1 to: 5177

477 LeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAs 493
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3 TTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAAGCTCTCAA 52

493 nGluMetTrpLysCysGlnAsnLeuArgHisGlnValLysAspLeuL 510
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
53 CGAATGTGGAAGTGTACAGACATGCTTCGGAGCCCATGTACCGAAGCTAT 102

510 euAspLeuLeuLysGlnProLysThrAspAlaSerValLysAlaLeuPhe 526
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
103 TGGATTTCACACAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTT 152

527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGl 543
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
153 GGAAACTGATGACCATAGCAAGAAATTTGCCTGACCCCGGGAAGCAC 202

543 nAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleA 560
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
203 AGATTTTGAAGAAATTTAAACAGGTTCTCGCGCATGATGAGAATCTC 252

560 rGlyGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
253 GCTCTCAGTTGAGTATTATTAATAGCCCAACCTGTTCTTTGCAACAA 302

577 GluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnPr 593
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
303 GATATTTGTGTGAGAAATATACCCGGAACTTGCAAAATCTCTTAAGCA 352

593 ofThrAsnProPheLeuGluMetIleLysPheLeuGluArgIleAlaP 610
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
353 ACAAAATCTTTCTAGAGATGTCATAATTTCTGTGGAAAGAAATCGCAC 402

610 roValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
403 CTGTGCACATTGATTCAGAGCAATAGTGCACACTAGTGAATGATGAAT 452

```

```

627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAs 643
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
453 AAGTCAATAGAGGGACAGCAGATGATGAAGAGGGGTGAAGTCCAGA 502

643 pGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrH 660
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
503 TACAGCTATCCGTTCCAGGACTTGAACCTTCTTAAGGTTCTGTCTTTAC 552

660 IspProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
553 ATCTTACCTCGTTCCACTCTGCAGAGACATATAGTCTTGTTCACAGTC 602

677 LeuLysMetAspAspGluLysValAlaGluAlaLeuGluGlnIlePhe 693
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
603 CTAAGAATGGAGGATGACAAAGGTAGCAGAGCTGTATTCAAAATTTTAG 652

693 sAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAlaL 710
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
653 AAATACAGGTCACAAATAGAAACAGACCTTCCCAGATACATCGACCT 702

710 euLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
703 TAATTCCCATTTTACATCAAAAGCAAGAGGGGTACTCCACACCAAGCA 752

727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGl 743
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
753 AAACAGGCTGTGCTATACACGCCCATATATTCACAAATAAAGAGTCCA 802

743 nPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnL 760
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
803 GCTTCACAGATTTTGGACCCACTCAGTAGGAGTCTGAATGCTGATGTC 852

760 euGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
853 CAGAACAACTTATACTCCACTAGTTTTCATGCGCCACATTTCTATGTTA 902

777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIl 793
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
903 GCACAGATCAGTTTGTCTCCCAATCTGTAAGTCTGTAAGTCTGTAAGT 952

793 eValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrL 810
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
953 TGTGAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAGAAATGGAA 1002

810 ysLeuTrpValProAspGluGluValSerProGluThrMetValLysIle 826
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1003 AACTGTGTCTCCAGATGAAGAGGTTTCCCTCTGAAGTACTAGCAAAAGTA 1052

827 GlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnH 843
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1053 CAGGCAATTAACCTTCTGGTAAAGGTGGTGTGGGTATGAAAGAACCA 1102

843 sSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHis 860
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1103 GTCTAAATCTGCCAATTCACCTTCGGTTATATATCAGCGATGTGGTTA 1152

860 exAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1153 GTGAGGTGACCTGCAGAGCAAAAGAGGATCAGTAAATCTGTATGTCT 1202

877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluPr 893
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1203 CCCTTCCGATTAGCTGCTGGTAGTCCCAATGAAGCTTGTCTCAGGAACC 1252

893 oCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuA 910
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1253 TTGTTTACCATGAAATATATACCCAGAACAGTTTTCAGCTCTGTGCAC 1302

910 IaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeu 926
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1303 TTATTAATGATCAGTGTACCAAGTAAGGACAGATATTTGCTCAGAAAGCTG 1352

927 HisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCy 943

```



```
539 roGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluasp 555
|||||
553 CTGGTAAGGCTCAGGATTTTCATGAAGAAATTCACACAGGTTTGAAGAT 602
|||||
556 AspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSe 572
|||||
603 GATGAGAAATAAGAAAGCAGTTAGAAAGTACTTGTAGTCCAAACATGCTC 652
|||||
572 rCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyA 589
|||||
653 CTGCACAGCGCTGAAGGTTGTGCGCGTGAATAACTTAAGAAGTTGGGCA 702
|||||
589 snProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeu 605
|||||
703 ACCCAACAGCGCTACAATCCCTTCTCGAATGATCAAGTTCTCTTG 752
|||||
606 GluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuII 622
|||||
753 GAGAGGATAGCACCCTGTGCATAGATACCGAATCTATCAGTGCTCTTAT 802
|||||
622 eLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluaspGluG 639
|||||
803 TAAACAAGTGAACAATACTAATAGATGAACAGCAGATGATGAAGATCAGG 852
|||||
639 lyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLysVal 655
|||||
853 GTGTTCCAAGTCAAGCCATCAGACAGCGTCTTGAACCTGCTTAAGGTA 902
|||||
656 LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSe 672
|||||
903 CTCTCATTTACATCCCATCTCATTTTCATTTCTGCTGAAACATTTGAATC 952
|||||
672 rLeuLeuAlaCysLeuLysMetaspGluLysValAlaGluAlaAlaL 689
|||||
953 ATTAAGTGGCTGTCTGAAATGGATGATGAAGTGAAGTGAAGTGCAC 1002
|||||
689 euGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHis 705
|||||
1003 TACAAATTTTCAAAACACAGGAGCAAAATTTGAAGAGATTTTCCACAC 1052
|||||
706 IleArgSerAlaLeuLeuProValLeuHisLysSerLysLysGlyPr 722
|||||
1053 ATCAGATCAGCGTTCGTCCTGTTTACATCACAATCTAAAAAGGACC 1102
|||||
722 oProArg GlnAlaLysTyAlaIleHisCysIleHisAlaIlePheSer 738
|||||
1103 CCCCCTCAGCCCAATATGCCATTCATTTGATCCATGCGATATTTTCT 1152
|||||
739 SerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
|||||
1153 AGTAAACAGACACAGTTTGCACAGATATTTGAGCCTCTGCATAGACCT 1202
|||||
755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyH 772
|||||
1203 AGATCCCAAGCAACCTGGAACATCTCATAACACCATTGCTTACTATTGCTC 1252
|||||
772 isIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrp 788
|||||
1253 ATATGCTCTCTTGCACCTGATCAATTTGCTGCTCTTTGAAATCTTG 1302
|||||
789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGl 805
|||||
1303 GTAGCTACTTTTCATTTGYGAAAGATCTTCTCATGAATGATCGGCTTCCAGG 1352
|||||
805 yLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluT 822
|||||
1353 GAAAAAGACAACTAACTTTGGGTTCCAGATGAAGAAAGTATCTCTCTGAGA 1402
|||||
822 hrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGly 838
|||||
1403 CAATGGTCAAAATTCAGGCTATTAAATGATGATGTTCCGATGCTACTTGGC 1452
|||||
839 MetLysAsnAsnHisSerLysSerGlyThrSerThrLeu ArgLeuLeuT 855
```

```
|||||
1453 ATGTAAMTTAMTCACAGTATATCAGGATCTTCTTCCTTAAGSATTTGCTAA 1502
|||||
855 hrThrIleLeuLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSer 871
|||||
1503 CACAATATATCGCATCGTGGGATTTGGMAGAMCAGGGGAAAT...TAG 1549
|||||
872 LysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVally 888
|||||
1550 TAACCAGATATGTCACGCTGAGACTTGTCTGGGAGTGTCTATTTRTGAA 1599
|||||
888 sLeuAlaGlnGluProCysTyrrHisGluIleIleThrLeuGluGlnTyrg 905
|||||
1600 GCTGGCACAAGAACCCCTGTTACCATGAATCTCACAATAGAACAAATATC 1649
|||||
905 InLeuCysAlaLeuAlaIleAsnAspGluCysTyrrGlnValArgGlnVal 921
|||||
1650 AGCTATGTCRTTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGTG 1699
|||||
922 Phe AlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuG 938
|||||
1700 TTTGGCCCAACAACCTTCGCAAGGCCCTTCCCGTTTCGGCTTCCACTTG 1749
|||||
938 luTyrrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 954
|||||
1750 AGAATATGGCAATCTGTGCGCTTTGTGCAAAAGATCTCTGTAAGGAGAGA 1799
|||||
955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgG 971
|||||
1800 AGAGCTCATGCTAGGCAATGTTTGGTGAATAATAAATGTAAGCGGGA 1849
|||||
971 u 971
:
1850 C 1850

seq_name: gb_pr1:AB014548
seq_documentation_block:
LOCUS AB014548 5177 bp mRNA PRI 06-FEB-1999
DEFINITION Homo sapiens mRNA for KIAA0648 protein, partial cds.
ACCESSION AB014548
VERSION AB014548.1 GI:3327109
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:HJ03994.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5177)
AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
JOURNAL Direct Submission
Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@info.kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
2 (sites)
REFERENCE Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
AUTHORS Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro
JOURNAL DNA Res 5 (3), 169-176 (1998)
MEDLINE
FEATURES
Location/Qualifiers
1..5177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HJ03994"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
```


1198 rAspLysArgAspSerAspLeuValArgSerGluLeuGluLysProA 1215
 |||||
 3503 TCACAGAGAGACGACTCGATCTTGAAGTCTGAATGGAGAGCCTA 3552
 1215 rGlyArgLysLysThrProValThrGluGlnGluLysLeuGlyMet 1231
 |||||
 3553 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3602
 1232 AspAspLeuThrLysLeuValGlnGlnLysProLysGlySerGlnAr 1248
 |||||
 3603 GATGACTTGACTAAGTGGTACAGACAGACAAACCTAAAGGAGTCAGCG 3652
 1248 gSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnInTrpP 1265
 |||||
 3653 AAGTCGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3702
 1265 rGluGluLysArgLeuLysGluAspLeuLeuGluAsnGluAspGluGln 1281
 |||||
 3703 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3752
 1282 AsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuG1 1298
 |||||
 3753 ANTATGCCCAAAAGGTTAAAGAGGCGCCACCAAAACCTCTTGG 3802
 1298 yGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlyS 1315
 |||||
 3803 TGGAGGTACACCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3852
 1315 erLysLysLysSerGlyProProAlaProGluGluGluGluGluGlu 1331
 |||||
 3853 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3902
 1332 ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgVa 1348
 |||||
 3903 AGACAAAGTGGAATACGGAACAGAGTCCAAAGACCAACACCGAGT 3952
 1348 lSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleG 1365
 |||||
 3953 GTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4002
 1365 lSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThr 1381
 |||||
 4003 ATCCACAGAGTCCACACAGAGGAGGAGGAGGAGGAGGAGGAGG 4052
 1382 ProSerProSerGlnProLysLysAsnVal 1391
 |||||
 4053 CCATCACCATCACACCAACCAACCAACCAACCAACCAACCAAC 4082

seq_name: gb_pr10:HSU50533

seq_documentation_block:
 LOCUS HSU50533 1852 bp mRNA PRI 27-NOV-1996
 DEFINITION Human BRCA2 region, mRNA sequence CG008.
 ACCESSION U50533
 VERSION U50533.1 GI:1531605
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1852)
 AUTHORS Couch, F.J., Rommens, J.M., Neuhausen, S.L., Belanger, C., Dumont, M.,
 Kenneth, A., Bell, R., Berry, S., Bogden, R., Cannon-Albright, L.,
 Farid, L., Frye, C., Hattier, T., Janek, T., Jiang, P., Kehrer, R.,
 Leblanc, J.-F., McArthur-Worrison, J., McSweeney, D., Miki, Y.,
 Peng, Y., Sanson, C., Schroeder, M., Snyder, S.C., Stringfellow, M.,
 Stroup, C., Swedlund, B., Swensen, J., Teng, D., Thakur, S., Tran, T.,
 Tranchesi, M., Welver-Feldhaus, J., Wong, A.K.C., Shizuya, H.,
 Labrie, F., Skolnick, M.H., Goldgar, D.E., Kamb, A., Weber, B.L.,
 Tavtigian, S.V. and Simard, J.
 Generation of an integrated transcription map of the BRCA2 region
 on chromosome 13q12-q13

JOURNAL Genomics 36 (1), 86-99 (1996)
 MEDLINE 96411650
 REFERENCE 2 (bases 1 to 1852)
 AUTHORS Simard, J.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular
 Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,
 Quebec City, Quebec G1V 4G2, Canada
 FEATURES
 Location/Qualifiers
 source
 1. 1852
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="13q12-q13"
 /note="CG008; DSEG numbers: D13S171 and D13S310"
 BASE COUNT 609 a 348 c 379 g 505 t
 ORIGIN
 11 others

alignment_scores:
 Quality: 2953.50 Length: 619
 Ratio: 4.947 Gaps: 4
 Percent Similarity: 96.446 Percent Identity: 95.153
 alignment_block:
 US-09-512-581-2 x HSU50533 ..
 Align seg 1/1 to: HSU50533 from: 1 to: 1852

356 LysValArgSerHisAspProGluAlaIleArgHisAspValIleVa 372
 |||||
 3 AAGGTGAGGTACATGACCCCTGAGGAAGCTATTAGACATGATGTTATTGT 52
 372 lSerIleValThrAlaAlaLysLysAspIleLeuValAsnAspHisL 389
 |||||
 53 GTCAATAGTACAGCTGCTAAAGAGGATATCTTCTGGTCAATGATCACT 102
 389 euLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArg 405
 |||||
 103 TACTTAAATTTGTGAGAGAGAGACATTAGACAAACGATGGAGAGTACGC 152
 406 LysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuCl 422
 |||||
 153 AAAGAGGCCATGATGGGACTTGCCTCAAAATTTATAAGAAATATGCTTTTACA 202
 422 nSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaIleTyrLysAspL 439
 |||||
 203 GTCACGAGCTGGAAGAGATGCTGCAAAACAGATAGCATGATCAAGACA 252
 439 ysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArgLeuLeuVal 455
 |||||
 253 AATTGCTACATATATATATCAAAATAGTATTGATGATCGACTACTTGT 302
 456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrCl 472
 |||||
 303 GAACGAGCTTTGCTCAATACATGCTTCTCACAATTTAGAAACATACAGA 352
 472 uArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlav 489
 |||||
 353 ACGGATGAATGCTTATATATCTTATCTGATGCCACACTGGATTTAAATGCTG 402
 489 alLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGln 505
 |||||
 403 TGAAAGCATTTGAATGAAATGTGAAATGTCAAAATCTGCTCGACATCAA 452
 506 ValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
 |||||
 453 GTAAAGGATTTGCTGACATGATTAAGCAACCCCAACAGATGCCAGTGT 502
 522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAsp 539
 |||||
 503 CAAGGCCATATTTTCAAAAGTGTGTTATTACAGAAATTTACCTGATC 552


```

|||||
1653 GAAATAACTAAGAAGTTGGCAACCCCAACAGCCTACAAATCCTTTCCT 1702
598 uGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspT 615
1703 GGAATGATCAAGTTCTTGGAGAGGATAGACCTGGCAGATAGATA 1752
615 hrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGly 631
1753 CCGAATCTATCAGTGCTCTTAAACAGAGTGAACAAATCAATAGATGA 1802
632 ThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAl 648
1803 ACAGCAGATGATGAAGATGAGGGTGTTCACACTGATCAAGCCATCAGAGC 1852
648 aGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheH 665
1853 AGGTCCTGAAGTCTTAAAGGTACTCTCATTTACACATCCATCTCATTTTC 1902
665 isSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAsp 681
1903 ATTCTGCTGAAACATTTGAATCATTTACTGGCTTGTCTGAAATGGATGAT 1952
682 GluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLy 698
1953 GAAAAAGTAGCAGAAGCTGCACATACAAATTTTCAAAAAACACAGGAACAA 2002
698 sIleGluLysAspPheProHisIleArgSerAlaLeuLeuProValLeuH 715
2003 AAPTGAAGAGGATTTTCCACACATCAGATCAGCCTTGGCTTCCTGTTTAC 2052
715 isHisLysSerLysLysGlyProArgGlnAlaLysTyrAlaIleHis 731
2053 ATCACAAATCTAAAAGAGGACCCCGCTCAAGCCAAATATGCCATTCAT 2102
732 CysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePh 748
2103 TGTATCCATCGGATATTTCTAGTAAAGAGACCCAGTTTGCACAGATATT 2152
748 eGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleT 765
2153 TGAGCCCTCTGCATTAAGAGCCTTAGATCCAGCAACCTGGAACATCTCATAA 2202
765 hrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPhe 781
2203 CACATTTGGTTACTATTGGTCATATTGGCTCTCTTGCACCTGATCAATTT 2252
782 AlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLeuLe 798
2253 GCTGCTCCTTTGAAATCTTTGGTAGCTACTTTTCATTTGTGAAAGATCTCT 2302
798 uMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProA 815
2303 CATGAATGATCGCTTCCAGGGAAAAAGACAACAACTTTGGGTTCCAG 2352
815 spGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMet 831
2353 ATGAAGAAGTATCTCTGAGACAATGGTCAAAATTCAGGCTATTAAATG 2402
832 MetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyTh 848
2403 ATGGTTTCGATGGCTACTTGAATGAAAAATAATACAGATAAATCAGGAAC 2452
848 rSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuT 865
2453 TTCTACCTTAAAGATTGCTACACAATAATTATGCATAGTAGTGAGACTTGA 2502
865 hrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 881
2503 CAGACAGGGGAAATTAGTAAACCAAGATATGTCAGCTCTGAGACTTGCT 2552
882 AlaGlySerAlaIleValLysLeuAlaGlnGluProCystTyrHisGluIl 898
|||||

```

```

2553 GCTGGAGTGCTATTGTGAAGCTGGCACAGAACCCTGTACCATGAAAT 2602
898 erIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluC 915
2603 CATCACATTAGAACAAATATCAGCTATGTGCATTAGCTATCAACGATGAAT 2652
915 ystYrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSer 931
2653 GCTATCAAGTAAGACAAGTGTTCGCCAGAACTTCACAAAGGCCCTTCC 2702
932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLy 948
2703 CGTTTACGGCTTCACCTTGAGTATATGGCAATCTGTCCCTTTGTGCAAA 2752
948 sAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysA 965
2753 AGATCTGTAAAGGAGAGAGAGCTCATGCTAGGCAATGTTTGGTGAAA 2802
965 snIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu 981
2803 ATATAAATGTAAAGCGGAGTATCTGAAGCAGCATGCAGCTGTAGTGAA 2852
982 LysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLe 998
2853 AAATTTATGTCTCTTACCAGAGTATGTGTTCATATACAATTCACCT 2902
998 uLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuL 1015
2903 TTTGGCAGCATGACCCAGATTTATGTCAAGATACAGGATATTGAACACTTA 2952
1015 ysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLys 1031
2953 AAGATGTTAAAGAAATGCTTTGGTTGTTCTGGAATAATTAATGGCTAAA 3002
1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLy 1048
3003 AATGAAAAAACAAGTCACGCTTTTATCAGAAAGATGTAGAAAAATATTA 3052
1048 sGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysL 1065
3053 ACAACAAAAGATGCCCAAGGACCAGATGATCAAAAATGAATGAAAAAC 3102
1065 euTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThr 1081
3103 TGTACACTGTGTGTGATGTGGCATGAAATATCATCATGTCAAAGAGTACT 3152
1082 ThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePh 1098
3153 ACATACAGTTTGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTT 3202
1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProG 1115
3203 CACTCAACCTGCACAAAGATTTTCAAGCAACCAAAAATTTATCTGCCTCCTG 3252
1115 luMetLysSerPheThrProGlyLysProLysThrThrAsnValLeu 1131
3253 AAATGAATCATTTTTTCACCTCGGAAAAACCTTAAACACCAATGTCTA 3302
1132 GlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLy 1148
3303 GGAGCTGTTTAAACAAGCCACTTTTCATCAGCAGGCAAGCAATCTCAGACCA 3352
1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProS 1165
3353 ATCATCAGCAATGGAACCTGTAAAGCAATGCAAGCAGCAGCTCAAAATCCA 3402
1165 erSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHis 1181
3403 GCTCTCTGGAGAATAAAGGGAGGCTTGATAGTTCTGAAATGGATCAC 3452
1182 SerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSe 1198
3453 AGTGAATAATGAAGATTACACAATGTCTTCACCTTTGCCGGGAAAAAAG 3502

```


ORIGIN

alignment_scores:
Quality: 6994.00 Length: 1360
Ratio: 5.154 Gaps: 0
Percent Similarity: 99.779 Percent Identity: 99.706

alignment_block:
US-09-512-581-2 x AB023196 ..

Align seg 1/1 to: AB023196 from: 1 to: 5110

32 ValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAs 48
|||||
3 CTTGAGACGATTAAAGATGCTTGTGAAACCTTTTATGGATATGGACCAAGCA 52
48 pSerGluGluGluLysGluLeuLysGluLeuLysGluLeuLysGluLeuLys 65
|||||
53 CTTGGAAGAAAGAAAGAGCTTTATTTAAACCTAGCTTTTACATCTTTGCTT 102
65 exAspPheLeuLysHisProGlyLysAspValArgLeuLeuValAla 81
|||||
103 CAGATTTTTTCTCAGCATCTCTGATAAGATGTTTCGCTTACTGGTAGCC 152
82 CysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrTh 98
|||||
153 TGCTGCTTGTGATATTTTACGATTTATGCTCTGAAAGCTCTTTACAC 202
98 rSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuL 115
|||||
203 ATCCCTGATAAACTAAAGGATATATTTATGTTTATAAAGACAGCTTGA 252
115 ysglyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeu 131
|||||
253 AGGGCTAGAGGATACAAAGACCCACAGTTCAATAGGATATTTTATTTA 302
132 LeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuG 148
|||||
303 CTTGAGAACATTTGCTGGGTCAAGTCATATACATATGCTTTGAGTTAG 352
148 uaspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerVal 165
|||||
353 AGATAGCAATGAAATTTTCCACCGAGTATACAGAACCTTTATTTTCAG 402
165 leAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMet 181
|||||
403 TAAACAATGCCACATCAAGAAGTCCATATGCACATGGTAGACCTTATG 452
182 SerSerIleileCysGluGlyAspThrValSerGlnGluLeuLeuAspTh 198
|||||
453 AGCTCTATTATTGTGAAGTGATACAGTGTCTCAGGAGCTTTTGGATAC 502
198 rValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAla 215
503 GGTTTTAGTAAATCTGGTACCTGCTCATAGAATTTTAAACAAGCAAGCAT 552
215 yAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluPro 231
|||||
553 ATGATTTGGCAAGGCTTTTACTGAAGAGGACAGCTCAAGCTATTGAGCCA 602
232 TyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIl 248
|||||
603 TATATTACCAATTTTTTAAATCAGGTTCTGATGCTTGGGAAACATCTAT 652
248 eSerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnI 265
|||||
653 CAGCGATTTGTCAGAGCATGCTCTTGACTTAATTTTGGAGCTCTACAATA 702
265 leAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeu 281
|||||
703 TTGATAGTCTATTTGCTGCTCTCTGTTTTACCCAGCTTGAATTTAAATTA 752

282 LysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
|||||
753 AAGAGCAATGATAATGAGGAGCGCTTACAAGTTGTTAAACTACTGCGAAA 802
298 smetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuT 315
|||||
803 AATGTTTGGGCAAGGATTCAGAAATTTGCTTCTCAAAACAAGCCACTTT 852
315 rpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeu 331
|||||
853 GGCAGTGCTACTTGGCAGGTTTAAATGATATCCATGTACCAATCCGCTG 902
332 GluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAl 348
|||||
903 GAATGTGTGAAATTTGCTAGCCATTTCTCATCAACCATCTGATTTAGC 952
348 alyAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluA 365
|||||
953 AAAAGACTTAAACAGAGTATCTTAAAGTGAGGTACATGACCTGAGGAAG 1002
365 laileArgHisAspValIleValSerIleValThrAlaAlaLysLysAsp 381
|||||
1003 CTATTAGACATGATGTTATTTGCTCAATAGTTACAGCTGCTAAAGAGAT 1052
382 IleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLe 398
|||||
1053 ATTCTTCTGCTCAATGATCATTACTTAAATTTTGTGAGAGAGAAACATT 1102
398 uaspLysArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnI 415
|||||
1103 AGACAAACGATGGAGAGTACGCAAGAGGCCATGATGGGACTTGGCCAAA 1152
415 letYrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLys 431
|||||
1153 TTTTAAAGAAATATGCTTTACAGTCAGCAGCTGGAAGAGATGCTGCAAAA 1202
432 GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSe 448
|||||
1203 CAGATAGCATGGATCAAGACAAATTTGCTACATATATATATCAAAATAG 1252
448 rIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValP 465
|||||
1253 TATTGATGATCGACTACTTGTGACGGATCTTTGCTCAATACATGGTTC 1302
465 roHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyr 481
|||||
1303 CTCACAATTTAGAACTACAGAACGGATGAATGCTTATATTACTTGTAT 1352
482 AlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCy 498
|||||
1353 GCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAATG 1402
498 sGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuLeuIleLysG 515
|||||
1403 TCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAAG 1452
515 InProLysThrAspAlaSerValLysAlaIlePheSerLysValMetVal 531
|||||
1453 ACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTT 1502
532 IleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLy 548
|||||
1503 ATTACAAGAAATTTACCTGATCTCTGTAAGGCTCAGGATTTTCATGAAGA 1552
548 sPheThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGluV 565
|||||
1553 ATTCACACAGGTTTAGAAGATGATGAGAAATAAGAAAGACAGCTTAGAAG 1602
565 alileuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArg 581
|||||
1603 TACTGTTTGTAGTCCACATGCTCTCTGCAAGCAGCTGAAAGTTGTGTGCGT 1652
582 GluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLe 598


```

3316 GTTTGGAACTCTCTAAAGACCGGTACTACAGCTCGTTTCTTCACRCAA 3365
1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLys 1117
|||||
3366 CCTGACAGAAGATTCAGTAACACCAAAATATATCTGCTCTCTGGAATGAA 3415
|||||
1117 sSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlav 1134
|||||
3416 ATCATTTTTCACCTCTGGAACACTAAACCAACCAATGTTCTAGGAGCTG 3465
|||||
1134 alAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer 1150
|||||
3466 TTAACAAGCCACTTTCATCAGCAGCAAGCAATCTCAGACCAAAATCATCA 3515
|||||
1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerPr 1167
|||||
3516 CGAATGGAACACTGAGGCAATGCAAGCAGCAGCTCAATCAATCAAGCTTCC 3565
|||||
1167 oGlyArgTleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA 1184
|||||
3566 TGGAGAATAAAGGGAGGCTTGATAGTTCTGAAATGGATCAGACAGTGA 3615
|||||
1184 snGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
|||||
3616 ATGAAGATTACACAATGCTTTCACCTTTGCCGGGAAAGAGTACAAG 3665
|||||
1201 ArgAspSerSerAspLeuValArgSerGluLeuGluLysProArgGlyAr 1217
|||||
3666 AGAGCAGCAGCTGATCTGTGTAAGCTCTGAATTTGGAGAAGCCTAGAGCAG 3715
|||||
1217 qLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspL 1234
|||||
3716 AAAAAAAGCCCTCAGACNACGGGAGGAGAAATAGGTATGATGACT 3765
|||||
1234 euThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArg 1250
|||||
3766 TGACTAAGTTGGTACAGAACAGACAGACTTAAAGGCAGTCAGCGAAGTCGG 3815
|||||
1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluG 1267
|||||
3816 AAAAGAGCCATACGGCTTCAGAAATCTGATGACAGCAGTGCGCTGAGGA 3865
|||||
1267 uLysArgLeuLysGluAspLleLeuGluAsnGluAspGluGlnAsnSerP 1284
|||||
3866 AAAGAGCTCAAGAACAGATATATTAGAAATGAAGATGAACAGATAGTC 3915
|||||
1284 roProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
|||||
3916 CGCCAAAAAGGGTAAAAGAGCGCGACCAACCAAAACCTCTTGGTGGAGGT 3965
|||||
1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy 1317
|||||
3966 ACACCAAAAGAGGAGCCCAACCAATGAAACTTCTAAAAAGAGGAGCAAAA 4015
|||||
1317 sLysSerGlyProProAlaProGluGluGluGluGluGluArgGlns 1334
|||||
4016 AAATCTGGACCTCCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4065
|||||
1334 erGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArg 1350
|||||
4066 GTGGAATATACGAACAGACAGAAAGTCCAAAAAGCAACAGCAGCGAGTGTCAAGG 4115
|||||
1351 ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh 1367
|||||
4116 AGAGCAGCAGAGAGAGAGAAATCTTCAATCTAGTCAATTTGAATCCAC 4165
|||||
1367 rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP 1384
|||||
4166 ACAGTCCACACACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4215
|||||
1384 roSerGlnProLysLysAsnVal 1391
|||||
4216 CATCACACCAAAAAAATGTG 4238

```

seq_name: gb_prl:AB023196

seq_documentation_block:

LOCUS Homo sapiens mRNA for KIAA0979 protein, partial cds.
 AB023196
 VERSION AB023196.1 GI:4589601

KEYWORDS
 SOURCE

Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
 SK plus clone:hj07056.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
1516 TGGATTAAATGCTGGAAGCATTTGAATGAAATGTGGAATGTCAAAAT 1565
501 LeuLeuArgHisGlnValLysAspLeuLeuLeuLeuLeuLeuLeuLeu 517
1566 GTGCTCCGACATCAAGTAAGGATTTGCTGACTTGAATTAAGCAACCCAA 1615
517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534
1616 AACAGATGCCAGTGTCAAGCCCATATTTCAAAAGTGATGGTTATTACAA 1665
534 rGAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550
1666 GAAGTTTACCTGATCTGTGAAGGCTCAGGATTTTCATGAAGAAATTCACA 1715
551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567
1716 CAGGTGTTAGAAGATGATGAGAAATTAAGAAAGCAGTTAGAAGTACTTGT 1765
567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584
1766 TAGTCCAACATGCTCTCAAGCAGGCTGAAGGTTGTGTGCGTGAATAA 1815
584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
1816 CTAAGAAGTTGGCAACCCCAACAGCCCTACAAATCTTTCTCGTGAATG 1865
601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrCluSe 617
1866 ATCAAGTTTCTCTTGAGAGGATAGCACCCTGTGCACATAGATACCGAATC 1915
617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634
1916 TATCAGTGTCTTATTAAACAAGTGAACAATCAATAGATGAACAGCAG 1965
634 sPaspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
1966 ATGATGAAGATGAGGTGTTCCAACTGATCAAGCCATCAGACAGGTCTT 2015
651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667
2016 GAACGTCTTAAGTACTCTCATTTACACATCCCATCTCATTTCAATCTGC 2065
667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysV 684
2066 TGAACATTTGAATCATTTACTGGCCCTGTCTGAAATGGATGATGAATAAG 2115
684 aAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
2116 TAGCAGAAGCTGCATACAAAATTTTCAAAAACACAGGAAGCAAAATGAA 2165
701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIle 717
2166 GAGGATTTTCCACATCATGATCAGCTTGTCTCTGTTTTATCATCACAA 2215
717 sSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleH 734
2216 ATCTAAAAAAGGACCCCGCTCAGCCCAATATGCCATTTCATTGTATCC 2265
734 isAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750
2266 ATCCGATATTTCTAGTAAGAGGCCAGCTTTGACAGATATTTGAGCCT 2315
751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767
2316 CTGCATAAGAGCCTAGATCCAGCAACCTGGAAACATCTCATACACCAAT 2365
767 uValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaP 784
2366 GGTACTATTGGTCATATTGCTCTCTCCTTGCACCTGATCAATTTGCTGCTC 2415
784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800

2416 CTTTGAATCTTTGGTAGCTACTTTTCATTTGAAAGATCTTCTCATGAAT 2465
801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluG 817
2466 GATCGCTTCCAGGAAAGAAACAATAAANCTTTGGGTTCCAGATGAAGA 2515
817 uValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834
2516 ACTATCTCTGAGACAAATGGTCAAAATTCAGGCTATTAAATGATGTTTC 2565
834 rGTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850
2566 GATGCTACTTGGAAATGAAAAATAATACAGTAAATCAGGAATCTTCTACC 2615
851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluG 867
2616 TTAAATGCTTAACAACAATATTTGATAGTAGGAGACTTCACAGAACA 2665
867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlys 884
2666 GGGCAAAATTTAGTAAACCAAGATATGTACGCTCTGAGACTTGTCTGGGA 2715
884 erAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
2716 GTGCTATTGTGAAGCTGGCACAAGAACCCCTGTACCATGAAATCATCA 2765
901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCystyrG 917
2766 TTAAACAATATCAGCTATGCTGATAGCTATCAACGATGAATGCTATCA 2815
917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
2816 ACTAAGACAAGTGTTCGCCAGAAACTTCACAAAGGCCCTTCCCGTTTAC 2865
934 rGLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950
2866 GCGTCTCCACTTGAGTATATGGAATCTGTGCCCTTTGTGCAAAAGATCCT 2915
951 ValLysGluArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
2916 GTAAAGGAGAGAAGAGCTCATCTAGGCAATCTTTGGTCAAAAATATAA 2965
967 nValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuL 984
2966 TGTAAAGCGGGAGTATCTCAAGCAGCATGCTGTAGTGTAGTAAAAATAT 3015
984 euSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
3016 TGTCTCTTCTACAGAGTATGTTTCCATATACAAATTCACCTTTTGGCA 3065
1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVa 1017
3066 CATGACCCAGATATTGTCAAGTACAGGATATTGAACAACCTTAAACATCT 3115
1017 lLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluA 1034
3116 TAAACAATGCTCTTGGTTTGTCTGGAATATTAATGGCTAAAAATGAAA 3165
1034 snAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr 1050
3166 ATAACAGTCACGCTTTTATCAGAAAGATGTTAGAAATATTAAAGCAACA 3215
1051 LysAspAlaGlnGlyProAspAlaLysMetAsnGluLysLeuIleThr 1067
3216 AAAGATGCCCAAGGACCATGATGCAAAAATGAATGAAAACTGTATACAC 3265
1067 rValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrS 1084
3266 TGTGTGTGATGTTGTCATGAAATATCATCATGTCAAGAGTACTATACATA 3315
1084 erLeuLeuSerProLysAspProValLeuProAlaArgPhePheThrGln 1100

/number=29
 3438..3583
 /gene="AS3"
 /number=30
 3584..3689
 /gene="AS3"
 /number=31
 3690..4129
 /gene="AS3"
 /number=32
 4130..4354
 /gene="AS3"
 /number=33
 4355..4810
 /gene="AS3"
 /number=34

BASE COUNT 1755 a 944 c 1074 g 1480 t
 ORIGIN

alignment_scores:

Quality: 7101.00 Length: 1391
 Ratio: 5.131 Gaps: 0
 Percent Similarity: 99.497 Percent Identity: 98.994

alignment_block:

US-09-512-581-2 x HSU95825

Align seg 1/1 to: HSU95825 from: 1 to: 5253

1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProPr 17
 66 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAATACATATCGCC 115
 17 oGlyValLysGluIleSerAspLysIleSerLysGluMetValArgA 34
 116 TGGGGTCAAGGAATATPCAGATAAAATATCTAAAGAGGAGATGGTGAG 165
 34 rGLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
 166 GATTAAAGATGGTGTGAAACCTTTTATGGATATGGACAGGACTCTGAA 215
 51 GluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67
 216 GAAGAAAGAGAGCTTTATTTAAACCTAGCTTTACATCTTGGCTTCAGATT 265
 67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84
 266 TTTTCTCAAGCATCTCGATAAAGATGTCGCTTACTGGTAGCCTGCTGCC 315
 84 euAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
 316 TTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCT 365
 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117
 366 GATAAACTAAGGATATATTTATGTTTATAACAGACAGTTGAAGGGCT 415
 117 uGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluA 134
 416 AGAGGATACAAAGAGCCACAGTTCATAGTAGTATTTTATTTGCTTGAGA 465
 134 snIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSer 150
 466 ACATGCTGGTCAAGTCATATAACATATGCTTTGAGTTAGAGATAGC 515
 151 AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAs 167
 516 AATGAAATTTTACCAGCTATACAGAACCTTATTTTCACTTATAACAA 565
 167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184
 566 TGGCCAAATCAGAAAGTCCATATGCATGTCATGTTAGACCTTATGAGCTCTA 615

184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
 616 TTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTA 665
 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLe 217
 666 GTAATCTGGTACCTGCTCATAGAATTTAAACAAGCAAGCATATGATT 715
 217 uAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleT 234
 716 GGCAAGGCTTTACTGAAGAGCAGCTCAAGCTATTGAGCCATATATTA 765
 234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250
 766 CCAATTTTAAATCAGTCTGATGCTTGGGAAACATCTATCAGCGAT 815
 251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSe 267
 816 TTGTCAGAGCATGTCTGACTTAATTTTGGAGCTCTACAATATTGATAG 865
 267 rHisLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284
 866 TCATTTGCTGCTCTGTTTTACCCAGCTTGAATTTAAATTAAGAGCA 915
 284 snAspAsnGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
 916 ATGATAATGAGGAGCGCTACAAGTTGTTAAACTACTGGGCAAAATGTT 965
 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317
 966 GGGCAAGGATTCAGAAATTTGGCTTCTCAAAACAAGCCACTTTGGCAGTG 1015
 317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334
 1016 CTACTTTGGGAGGTTTAAATGATATCCATGTACCATCCGCTGGAAATGTG 1065
 334 allysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350
 1066 TGAATTTGCTAGCCATTGCTCATGAACCTCTGATTTAGCAAAAGAC 1115
 351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
 1116 TTAACAGAGTATCTTAAAGTGAGGTACATGACCTCGAGGAAGCTATTAG 1165
 367 gHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
 1166 ACATGATGTTATTGTGTCAATAGTTACAGCTGTCTAAAAGGATATTCTTC 1215
 384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
 1216 TGGTCAATGATCACTTACTTAATTTTGGGAGAGAGAACATTAGACAAA 1265
 401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417
 1266 CGATGGAGAGTACGCAAGAGCCATGATGGACTTGCCTCAAAATTTATAA 1315
 417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434
 1316 GAAATATGCTTTTACAGTCAGCAGCTGGAAAGATGCTGCAAAACAGATAG 1365
 434 latrPileLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450
 1366 CATGGATCAAGACAAATTTGCTACATATATATATCAAAATAGATTGAT 1415
 451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467
 1416 GATCGACTACTTGTGGAACGGATCTTTTGCCTCAATACATGTTCTCTCACA 1465
 467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuThrAlaThrL 484
 1466 TTTAGAAATACAGAACGGATGAATGCTTATATATTACTTGTATGCCACAC 1515


```
884 erAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleThr 900
|||||
2777 GTGCTATTGTGAAGCTGGCAAGAACCCCTGTTACCATGAATCATACA 2826

901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrG1 917
|||||
2827 TTAGAACAATATCATCACTATGCTATGCTATCATCAACGATGAATGCTATCA 2876

917 nValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuA 934
|||||
2877 AGTAAGACAAGTGTTCGCCAGAACCTTCACAAAGGCCCTTTCCCGTTTAC 2926

934 rgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspPro 950
|||||
2927 GGCCTCCACTTGAGTATATGCAATCTGTGCCCTTTGTGCAAAAGATCCT 2976

951 ValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAs 967
|||||
2977 GTAAAGGAGAGAGAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAA 3026

967 nValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuL 984
|||||
3027 TGTAAAGCGGGAGTATCTGAAGCAGCATGCTGTTAGTGAAAAATATAT 3076

984 euSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
|||||
3077 TGTCTCTTACACAGATGATGTTGTTCCATATACAAATTCACCTTTTGCCA 3126

1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVa 1017
|||||
3127 CATGACCCAGATTATGTCAAAGTACAGGATATTCAACAACCTTAAAGATGT 3176

1017 LysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLysAsnGluA 1034
|||||
3177 TAAAGAATGCTTTGGTGTGTTCTGGAAATATTAATGGCTAAAAATGAAA 3226

1034 snAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThr 1050
|||||
3227 ATAACAGTCAGCCTTTATCAGAAGATGGTAGAAATATATAACAACA 3276

1051 LysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrTh 1067
|||||
3277 AAAGATGCCAAGGACCAAGATGATGCAAAATGAATGAAAACTGTACAC 3326

1067 rValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrS 1084
|||||
3327 TGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACA 3376

1084 erLeuGluSerProLysAspProValLeuProAlaAaGpPheThrGln 1100
|||||
3377 GTTTGGAATCTCCTTAAGACCCGGTACTACCAGCTCGTTCTTCACATCAA 3426

1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMetLy 1117
|||||
3427 CCTCACAAAGAAATTCAGTAACACCAAAATATATCTGCTCTCGTGAATGAA 3476

1117 sSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlav 1134
|||||
3477 ATCAATTTTCTACTCCTGGAAACCTAAACCAACCAATGTTCTTAGGAGCTG 3526

1134 alAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSer 1150
|||||
3527 TTAACAAGCCACTTTCATCAGCAGCAAGCAATCTCAGACCAATCATCA 3576

1151 ArgMetGluThrValSerAsnAlaSerSerSerSerSerAsnProSerSerPr 1167
|||||
3577 CGAATGGAATGTTAAGCAATGCAAGCAGCAGCTCAATCAATCAAGCTCTCC 3626

1167 oglyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluA 1184
|||||
3627 TGGAAAGATAAAGGGAGGCTTGATGTTCTGAATGATCAGTCAAGTAAAA 3676

1184 snGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
```

```
|||||
3677 ATGAAGATTACACAATGCTTTCACCTTTGCCGGGAAAAAAGTGACAAG 3726

1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyAr 1217
|||||
3727 AGAGCAGACTCTGATCTTGTAAAGCTCTGAATTTGGAGAAGCCCTAGAGCCAG 3776

1217 qLysLysThrProValThrGluGlnGluLysLeuGlyMetAspAspL 1234
|||||
3777 GAAAAAACGCCCTGCAGAACAGGAGGAGAAATTAGGTATGATGACT 3826

1234 euThrLysLeuValGlnGlnLysProLysGlySerGlnArgSerArg 1250
|||||
3827 TGACTAAGTTGGTACAGAACAGAAACCTTAAAGGCAGTCACCGAAGTCGG 3876

1251 LysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluG1 1267
|||||
3877 AAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGA 3926

1267 uLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerP 1284
|||||
3927 AAAGAGCTCAAGAAGATATATTAGAAAATGAAGATGAACAGAAATAGTC 3976

1284 roProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
|||||
3977 CGCCAAAAAGGGTAAAGAGCGCCGACCACCAAAACCTCTTGGTGGAGGT 4026

1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLy 1317
|||||
4027 ACACCAAAAGAAGACCAACAATGAAACTTCTAAAAAAGGAAGCAAAA 4076

1317 sLysSerGlyProProAlaProGluGluGluGluGluGluArgGlnS 1334
|||||
4077 AAAATCTGGACCTCCAGCACAGGAGGAGGAGGAAGAAAGAAAGACAAA 4126

1334 erGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArg 1350
|||||
4127 GTGCAATACGGAACAGAAAGTCCAAAAGCAACAGACCGAGTGTCAAGG 4176

1351 ArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerTh 1367
|||||
4177 AGAGCAGACAGAGAGAGAGATCTCTCTGAATCTAGTCAATTTGAATCCAC 4226

1367 rGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerP 1384
|||||
4227 ACAGTCACACACCACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4276

1384 roSerGlnProLysLysAsnVal 1391
|||||
4277 CATCACAAACCAAAAAAATGTC 4299
```

seq_name: gb_pr10:HSU95825

```
seq_documentation_block:
LOCUS      HSU95825      5253 bp      mRNA      PRI      05-APR-1999
DEFINITION Human androgen-induced prostate proliferative shutoff associated
            protein (AS3) mRNA, complete cds.
ACCESSION  U95825
VERSION    U95825.2  GI:4559409
KEYWORDS   human.
```

```
ORGANISM   human.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 5253)
AUTHORS   Geck,P., Szelel,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.
TITLE     Androgen-induced proliferative shutoff in prostate cancer cells
JOURNAL   Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
REFERENCE  2 (bases 1 to 5253)
AUTHORS   Geck,P., Szelel,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
TITLE     Early gene expression during androgen-induced inhibition of
            proliferation of prostate cancer cells: a new suppressor candidate
            on chromosome 13, in the BRCA2-Rb1 locus
```


927 TCATTGCTGCTCTCTGTTTACCACGAGCTTGAATTTAAATTAAGAGCA 976
284 sAspAsnGluGluArgLeuGlnValValLysLeuAlaLysMetPhe 300
977 ATGATAATGAGGAGCCCTACAAAGTTGTTAACTACTGCAAAATGTTT 1026
301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCy 317
1027 GGGGCAAGGATTCAAGATTGCTTCTCAAAACAAGCCACTTTGCGACGTG 1076
317 sTrpLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysV 334
1077 CTACTTGGCAGGTTTAATGATATCCATGTACCAATCCGCCCTGGAAATG 1126
334 aLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350
1127 TGAATTTGCTAGCCATTGCTCATGAACCATCTGATTTAGCAAAAGAC 1176
351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleAr 367
1177 TTAACAGAGTATCTTAAAGTGAGGTACATGACCTCGAGGAAGCTATTAG 1226
367 qHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuL 384
1227 ACATGATGTTATTTGTCTCAATAGTTACAGCTGCTAAAAGGATATCTTC 1276
384 euValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
1277 TGGTCAATGATCATTACTTAATTTTGTGAGAGAGAACATTAGACAAA 1326
401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLy 417
1327 CGATGAGAGTAGCGCAAAAGAACCCATGATGGGACTTGGCCAAATTTATA 1376
417 sLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleA 434
1377 GAAATATGCTTTACAGTCAGCAGCTGGAAGAGATGCTGCAAAACAGATAG 1426
434 laTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAsp 450
1427 CATGGATCAAGACAAAATGGCTACATATATATATCAAAATAGTATTGAT 1476
451 AspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAs 467
1477 GATCGACTACTTGTGAAGCGGATCTTGTCTCAATACATGGTTCCCTCACAA 1526
467 nLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrL 484
1527 TTTAGAACTACAGAAGGATGAATGCTTATATTTACTTGTATGCCACAC 1576
484 euAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
1577 TGGATTTAAATGCTGTGAAGGATTTGAATGAATGTGAAGATGTCAAAAT 1626
501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLysGlnProly 517
1627 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTTAAGCAACCCAA 1676
517 sThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrA 534
1677 AACAGATGCCAGTGTCAAGCCATATTTCAAAAGTGATGGTTATTACAA 1726
534 rGAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550
1727 GAAATTTACCTGATCTGTAAGGCTCAGGATTTTCATGAAGAAATTCACA 1776
551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVa 567
1777 CAGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGCTTAGAAGTACTTGT 1826
567 lSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleT 584
1827 TAGTCCAACATGCTCTGCAACGAGGCTGAAGGTTGTGCGCTGAATAA 1876

584 hrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
1877 CTAACAAGTTGGCAACCCCAACAGCCTTACAAATCCTTTCTCTGGAATG 1926
601 lleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSe 617
1927 ATCAAGTTTCTCTGGAGAGGATAGCCTGTGCACATAGATACCGAATC 1976
617 rIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaA 634
1977 TATCAGTCTCTTATTAACAAGTGAACAATCAATAGTAGGAACAGCAG 2026
634 sPAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
2027 ATGATGAAGATGAGGTGTTCCAAGTGAATCAAGCCATCAGACGAGTCTT 2076
651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAl 667
2077 GAATGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCAATCTGC 2126
667 aGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspGluLysV 684
2127 TGAACATTTGAATCATTTACTGCTGCTGCTGAATGGATGATGAAAAAG 2176
684 aIaAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
2177 TAGCAGAAGCTGCATCAAAATTTTCAAAACACAGGAAGCAAAATTTGAA 2226
701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLy 717
2227 GAGGATTTTCCACACATCAGATCAGCCTTGTCTGCTTTTACATCACAA 2276
717 sSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleH 734
2277 ATCTAAAAAGGAGCCCGCTCAAGCCAAATATGCCATTCATTGTATCC 2326
734 lsAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluPro 750
2327 ATGCGATATTTTCTAGTAAAGAGAGCCAGTTTGCACAGATATTTGAGCCT 2376
751 LeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLe 767
2377 CTGCATAGAGCCTAGATCCAAGCAACCTGGAAACATCTCATACACCAAT 2426
767 vAlaThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlap 784
2427 GGTACTATTGCTCATATTGCTCTCTCTTGCACCTGATCAATTTGCTGCTC 2476
784 roTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
2477 CTTGAAATCTTTGGTAGCTACTTTCATTTGTGAAAGATCTTCTCATGAAT 2526
801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGl 817
2527 GATCGCTTCCAGGAAAAGCAACTAAACTTTGGTTCCAGATGAAGA 2576
817 vAlSerProGluThrMetValLysIleGlnAlaIleLysMetMetValA 834
2577 ACTATCTCTGAGACAATGGTCAAAATTCAGGCTATTAAATGATGGTTC 2626
834 rGrTrpLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThr 850
2627 GATGCTACTTGGAAATGAATAATAATACAGTAAATCAGGAACCTTCTACC 2676
851 LeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGl 867
2677 TTAAGATGCTAACACAATAATTTGCATGATGGAGACTTGCACAGAACA 2726
867 nGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlys 884
2727 GGGGAAATTTAGTAAACCAGATATGTACGCTGAGACTTGTGCTGGGA 2776


```
exon      3645..3750
           /number=31
exon      3751..4190
           /number=32
misc_feature
/note="complement(3864..4286)
/note="matches EST A1680124 from clone IMAGE:2264521"
misc_feature
/note="matches EST N83765 from clone KK2745"
misc_feature
/note="matches EST F12354 from clone c-38h11."
exon      4191..4298
           /number=33
misc_feature
/note="matches EST T64867 from clone 22264"
exon      4299..4434
           /number=34
misc_feature
/note="Join(4398..4533,4518..4660,4659..4675,4674..4693)
/note="matches EST H44781 from clone 188032"
exon      4435..7444
           /number=35
misc_feature
/note="matches EST A1681422 from clone IMAGE:2272632"
misc_feature
/note="matches EST AA911125 from clone IMAGE:1416862"
misc_feature
/note="matches EST A1275866 from clone IMAGE:1877482"
misc_feature
/note="matches EST AA987361 from clone IMAGE:1603350"
misc_feature
/note="matches EST A1299646 from clone IMAGE:1898010"
misc_feature
/note="matches EST AA447433 from clone 784555"
misc_feature
/note="matches EST H93424 from clone 220700"
misc_feature
/note="Join(4573..4649,4650..4815)
/note="matches EST W20070 from clone 306049"
misc_feature
/note="complement(4629..4937)
/note="matches EST AW089775 from clone IMAGE:2593146"
misc_feature
/note="matches EST AA747568 from clone IMAGE:1269696"
misc_feature
/note="complement(4803..5332)
/note="matches EST A1911784 from clone IMAGE:2329873"
misc_feature
/note="complement(4828..5296)
/note="matches EST AA563884 from clone IMAGE:1013990"
misc_feature
/note="4861..5243
/note="matches EST D20082 from clone pml283."
misc_feature
/note="complement(join(4872..4921,5281..5332))
/note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature
/note="Join(4872..4921,5281..5371)
/note="matches EST A1177347 from clone 621810"
misc_feature
/note="complement(4872..4921)
/note="matches EST AW048929 from clone
UI-M-BH1-amo-c-09-0-UI"
misc_feature
/note="4872..4921
/note="matches EST A1154932 from clone 599194"
misc_feature
/note="complement(4876..5333)
/note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature
/note="complement(join(4922..4975,5281..5332))
/note="matches EST A1012649 from clone RPLBA02"
misc_feature
/note="complement(4922..4975)
/note="matches EST AA899586 from clone
UI-R-E0-cy-e-02-0-UI; matches EST A1172172 from clone
RMUBV82"
misc_feature
/note="complement(4931..5332)
/note="matches EST AA625961 from clone 745478"
misc_feature
/note="complement(4951..5332)
/note="matches EST A1140790 from clone IMAGE:1566222"
misc_feature
/note="complement(5021..5309)
/note="matches EST H45693 from clone 188292"
```

alignment_scores:
Quality: 7155.00
Ratio: 5.155
Length: 1391
Gaps: 0

Percent Similarity: 99.784 Percent Identity: 99.712
alignment_block:
US-09-512-581-2 x HS26H231 ..
Align seg 1/1 to: HS26H231 from: 1 to: 7444
1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyProPr 17
127 ATGGCTCATTCAAAGACTAGACCAATGATGGAAAAATTACATATCGGCC 176
17 OGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgA 34
177 TGGGGTCAAGGAATAATACAGATAAAATATCTAAAGAGAGAGATGAGAC 226
34 rGLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
227 GATTAAGATGGTGTGAAAACCTTTTATGGATATGGACGAGACTCTGAA 276
51 GluGluLysGluLeuTyLeuAsnLeuAlaLeuHisLeuAlaSerAspPh 67
277 GAAGAAAGAGAGCTTTATTTAAACCTAGCTTTACATCTTGCCTTCAGATT 326
67 ePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysL 84
327 TTTTCTCAAGCATCTCTGATAAAGATGTTCTGCTTACTGTAGCTGCTGCC 376
84 euAlaAspIlePheArgIleTyAlaProGluAlaProTyThrSerPro 100
377 TTGCTGATATTTTTCAGGATTTATGCTCTGAAAGCTCTTACACATCCCT 426
101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLe 117
427 GATAAACTAAAGGATATATTTATGTTTAAACAGACAGTTTGAAGGGCT 476
117 uGluAspThrLysSerProGlnPheAsnArgTyPheTyLeuLeuGluA 134
477 AGAGGATACAAAGAGCCCAATTCATAGTATTTTATTACTTACTGAGA 526
134 snIleAlaIleValLysSerTyAsnIleCysPheGluLeuGluAspSer 150
527 ACATTTGCTGGTCAAGTCATATAACATATGCTTTGAGTTAGAGATAGC 576
151 AsnGluIlePheThrGlnLeuTyArgThrLeuPheSerValIleAsnAs 167
577 AATGAAATTTTACCCAGCTATACAGAACCTTATTTTCAGTTATATAACA 626
167 nGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerI 184
627 TGGCCCAATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTA 676
184 leIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
677 TTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGTTTA 726
201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyAspLe 217
727 GTAATCTGGTACCTGCTCATAGAATTTAAACAAGCAAGCATATGATTT 776
217 uAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyIleT 234
777 GGCAAGGCTTTACTGAAGAGGAGAGCTCAAGCTATTGAGCCATATATTA 826
234 hrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAsp 250
827 CCAATTTTATATCAGGTTCTGATGCTTGGGAAAAACATCTATCAGCGAT 876
251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyAsnIleAspSe 267
877 TTGTACAGAGCATGCTTTGACTTAATTTTGAGGCTCTACAAATATTGATG 926
267 rHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerA 284

gb_hgt19:AC090533 - 161.50 90.35 8.5e+03 241602 ! AC090533 Mus musculus strain
gb_hgt1:AC006842 - 161.50 88.46 1.1e+04 299015 ! AC006842 Caenorhabditis eleg
gb_in1:AE003433 - 161.50 88.33 1.1e+04 303411 ! AE003433 Drosophila melanoga
gb_pr10:HUM1GAA - 161.50 88.02 1.1e+04 314205 ! M36067 Human DNA ligase I mRNa
gb_pat2:AX072942 + 161.00 128.60 63.02 3083 ! AX072942 Sequence 53 from Pate
gb_pl3:DG4087 + 161.00 127.67 70.95 3422 ! AX072942 Sequence 53 from Pate
gb_ov:GGBRG1PRT + 161.00 126.22 85.51 4033 ! D64087 Daucus carota mRNA for
gb_in1:AF036760 + 161.00 124.50 106.52 4893 ! X91637 G-gallus mRNA for BRG1
gb_in2:AF009623 + 161.00 123.30 124.35 5607 ! AF036760 Rattus norvegicus BRG
gb_in2:AF013277 + 161.00 122.65 135.06 6030 ! AF009623 Parascaris univalens
gb_in2:AF013277 + 161.00 122.11 144.82 6412 ! AF013277 Bombyx mori topoisome

seq_name: gb_pr8:HS26H231

seq_documentation_block:

LOCUS HS26H231 7444 bp mRNA PRI 25-JAN-2000
DEFINITION Novel human gene mapping to chromosome 13.
ACCESSION AL137201
VERSION AL137201.1 GI:6759511

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 7444)

Rhodes,S. and Huckle,E.

Direct Submission

Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk

This cDNA sequence was assembled from public domain ESTs and single

pass sequencing reads from expressed DNA templates, aligned to the

genomic DNA sequence from the bacterial clones 26H23 (284467),

267P19 (275889) and 49J10 (284572).

The EST sequences listed match this sequence with an identity of at

least 95% between the coordinates shown.

Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene

Sanger Centre name : 26H23.Cl3.1.

Location/Qualifiers

1. .7444

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="13"

/map="13q12.3"

1. .107

/number=1

108. .234

/number=2

127. .4470

/codon_start=1

/product="hypothetical protein"

/protein_id="CAB69911.1"

/db_xref="GI:6759512"

/translation="MAHSKTRTNDGKITPPGVKEISDKISKEEMVRLKMKVKTMD

MDQSEEEKLYNLALHSLASDFLKHDPKDVLLVACCLADIFRIYAPAPVTSPPK

LKDFEFTQLGLEDTSKQPNRYELLENATWYKSYNICELEDSNEIFQLYRT

LFVSINNGHQKVMHVRDLMSSICEGDTVSQELLDVLVNLVPAHKNKQAYDIA

KALKRTAQAEPTITNFQVLMGKTSISDLSEHFVDLILEYNDSDHSILSVLPQ

LEFKLSNDNEERLVYLLAKFMGKDSQNPQWYLGRENDIHVPLRECV

KFASHCLMNHDPDLAKDLEYLKVSRHDPPEAIRHDVIVSTAAKDLLLVNDHLLNF

VRETDLKRVRKEAMGLAOLYKVALQSAAGKAAKOIAWKDKLLHYIYQNSID

DRLLVERIPAQYVPHNLETETRMKCLYLYLYALDLNAVVALNEMKQNLRRQVND

LLDLIKQPTDASVKAIFSKVWITRNLPDPGQAFMKFTQVLEDEDEKIRKOLEVL

VSPSCQKAGCVREITKLGNPQPTNPFLEMIKFLERAPIRVHDTESIALIKQ

VNKSIDGTADEGVPTDQAIIRAGLELLKVLSTFTHPISFSAETFEELACLKMDDE

KVAEAAQIKFTKQTEEDPHIRSAALLVHLHKKKKPPQAKYAIHCHIAIFSSK

ETQPAQIFPELHSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFVKDL

MNRLPGKTTKLWPPDEVPSTWYKQIAIKMWVRLKMKNNHNSKSGTSLRLTIT

ILHSDGLTIGQKISRPDSRLRAGSAIVKLAQPCYHEITITLEQYQCALAINDE

CYQVQVFAQKHLGSLRSLPLEYMAICALCAPDVKERRAHARQCLVKNINVRREY

LKQAAVSEKLLSLLPEYVPTTHLAHDPDYVKQDIEQLKDVRECLFWLEILMA

KDNVSHAFIRKNVETIKQTKDAQGPDARKNEKLYTVCDVAMNIIMSKSTVTSLESP

KDPVLPARFTQDPKFNPKVLPPEMKSFPTPGKPTTNVLGAVNKPPLSSAGKOSQ

TKSSRMETVSNASSSSPSPGRIGKRLDSEMDHSENYDTWSSPLPGKSKDRDSD

FEATURES

source

exon

exon

CDS

DLVRSELEKPRGRKKTPTVTEQEEKLGMDLTKLVQEQPKGQSRKRKRGHTASDSDEQ
QWPEERKRLKILENEDEQNSPKKGRGRPPKPLGGGTPEEPTKTKSKGKKKSG
PPAPEEEREEERGSGNTEQKSKQHRVSRRAQQAESPSSAESTQSTPQKGRGRP
SKTSPSPQPKKNVVRGSRKQAATKENDSSEEDVDFQGSPPVDDIPQETEEEEVSTVN
VRRSAKRERR"
235. .438
/number=3
439. .525
/number=4
526. .623
/number=5
624. .750
/number=6
751. .831
/number=7
832. .972
/number=8
973. .1088
/number=9
1089. .1183
/number=10
1184. .1329
/number=11
1330. .1481
/number=12
1482. .1595
/number=13
1596. .1677
/number=14
1678. .1726
/number=15
1727. .1866
/number=16
1866. .2046
/note="matches EST AA687846 from clone IMAGE:1219676"
1866. 1941
/note="matches EST AA687776 from clone IMAGE:1219628"
/note="matches EST
1867. .1982
/number=17
1983. .2088
/number=18
2089. .2249
/number=19
2250. .2373
/number=20
2374. .2532
/number=21
2533. .2601
/number=22
2602. .2738
/number=23
2739. .2862
/number=24
2863. .3067
/number=25
complement(2912. .3302)
/note="matches EST AA630616 from clone 856164"
3068. .3182
/number=26
3183. .3315
/number=27
3302. .3572
/note="matches EST AA333118"
3316. .3435
/number=28
3439. .3818
/note="matches EST AA282190 from clone IMAGE:704631"
3436. .3498
/number=29
3499. .3644
/number=30
complement(3586. .4065)
/note="matches EST AI655429 from clone IMAGE:2239331"

| | | | | | | | |
|--------------------|---|--------|--------|---------|--------|---|---------------------------------|
| gb_inl:AE003489 | + | 168.00 | 93.48 | 5.7e+03 | 302915 | ! | AE003489 Drosophila melanog |
| gb_pr10:HUMSRP1570 | - | 167.50 | 137.23 | 20.82 | 2076 | ! | LI4076 Human pre-mRNA splicing |
| gb_pr9:HSAC001050 | + | 167.50 | 133.46 | 33.77 | 3177 | ! | AC001070 Homo sapiens (Subclon |
| gb_pl3:DB4225 | + | 167.50 | 129.15 | 58.67 | 5166 | ! | D84225 Arabidopsis thaliana fr |
| gb_ov:DB9991 | + | 167.50 | 127.92 | 68.77 | 5941 | ! | D89991 Cyprinus carpio mRNA f |
| gb_pr5:AF111784 | + | 167.50 | 127.89 | 68.97 | 5956 | ! | AF111784 Homo sapiens myosin I |
| gb_ov:XELTANABIN | + | 167.50 | 126.44 | 83.11 | 7019 | ! | M99387 xenopus laevis tanabin |
| gb_hg6:AC020364 | + | 167.50 | 124.74 | 103.31 | 8500 | ! | AC020364 drosophila melanogast |
| gb_pr8:HS1340191 | + | 167.50 | 122.34 | 140.55 | 11145 | ! | AL096858 Novel human gene map |
| gb_r01:AF150755 | + | 167.50 | 118.43 | 232.15 | 17333 | ! | AF150755 Mus musculus microtu |
| gb_ba1:AE000561 | - | 167.50 | 117.79 | 251.92 | 18629 | ! | AE000561 Helicobacter pylori |
| gb_hg14:AC068192 | - | 167.50 | 110.29 | 659.07 | 43417 | ! | AC068192 Homo sapiens clone F |
| gb_pl1:AB006570 | - | 167.50 | 104.21 | 1.4e+03 | 86263 | ! | AB006702 Arabidopsis thaliana |
| gb_pr8:HS134019 | - | 167.50 | 104.15 | 1.4e+03 | 86897 | ! | AL034555 Human DNA sequence |
| gb_hg21:ALJ355994 | - | 167.50 | 98.90 | 2.8e+03 | 17095 | ! | ALJ35599 Homo sapiens chromo |
| gb_hg11:AC0026478 | - | 167.50 | 97.37 | 3.5e+03 | 186827 | ! | AC026478 Mus musculus chromo |
| gb_inl:AC009257 | - | 167.50 | 95.98 | 4.1e+03 | 218565 | ! | AC009257 Drosophila melanog |
| gb_hg16:AC074307 | - | 167.50 | 95.90 | 4.2e+03 | 220469 | ! | AC074307 Mus musculus chromo |
| gb_hg17:AC079431 | + | 167.50 | 93.15 | 5.9e+03 | 300695 | ! | AC079431 Mus musculus chromo |
| gb_pr1:AB046029 | + | 167.50 | 136.76 | 22.13 | 2095 | ! | AB046029 Macaca fascicularis E |
| gb_ba2:AF305602 | + | 167.00 | 132.04 | 40.54 | 3569 | ! | AF305600 Borrelia burgdorferi |
| gb_ba2:AF305600 | + | 167.00 | 132.04 | 40.54 | 3569 | ! | AF305602 Borrelia burgdorferi |
| gb_ba2:AF305604 | + | 167.00 | 132.04 | 40.54 | 3569 | ! | AF305604 Borrelia burgdorferi |
| gb_ba2:AF305606 | + | 167.00 | 132.04 | 40.54 | 3569 | ! | AF305606 Borrelia burgdorferi |
| gb_ba2:AF305609 | + | 167.00 | 131.64 | 42.64 | 3731 | ! | AF305609 Borrelia burgdorferi |
| gb_pr10:HS040847 | + | 167.00 | 129.43 | 56.62 | 4789 | ! | U40847 Human Treacher Collins |
| gb_pr10:HS07366 | + | 167.00 | 129.34 | 57.31 | 4840 | ! | U76366 Human Treacher Collins |
| gb_ov:AF017250 | + | 167.00 | 128.19 | 66.42 | 5511 | ! | AF017250 Oreochromis aureus vi |
| gb_ba1:AE001131 | + | 167.00 | 121.15 | 163.70 | 12190 | ! | AE001131 Borrelia burgdorferi |
| gb_ba1:AE001153 | + | 167.00 | 118.17 | 240.00 | 17070 | ! | AE001153 Borrelia burgdorferi |
| gb_pl3:NC1586 | - | 167.00 | 106.77 | 1.0e+03 | 61843 | ! | AL353822 Neurospora crassa DN |
| gb_v12:HHU43400 | + | 167.00 | 99.22 | 2.7e+03 | 144861 | ! | U43400 Human herpesvirus-7 C |
| gb_hg5:AC016932 | + | 167.00 | 98.93 | 2.8e+03 | 149845 | ! | AC016932 Homo sapiens chromo |
| gb_v11:AF037218 | + | 167.00 | 98.74 | 2.9e+03 | 153080 | ! | AF037218 Human herpesvirus 7 |
| gb_hg10:AC0925326 | + | 167.00 | 98.05 | 3.2e+03 | 165490 | ! | AC0925326 Homo sapiens chromo |
| gb_pr4:AC090884 | + | 167.00 | 97.97 | 3.2e+03 | 166991 | ! | AC090884 Homo sapiens chromo |
| gb_hg3:AC012474 | - | 167.00 | 96.34 | 3.9e+03 | 200379 | ! | AC012474 Homo sapiens chromo |
| gb_hg16:AC073809 | + | 166.50 | 135.13 | 27.27 | 278632 | ! | AC073809 Mus musculus clone |
| gb_pr5:AK023416 | + | 166.50 | 133.35 | 34.27 | 2408 | ! | AK023416 Homo sapiens cDNA FLJ |
| gb_ov:BUHCY342 | + | 166.50 | 129.52 | 55.97 | 4534 | ! | AF091342 Bos taurus neurofilam |
| gb_ba2:BUHCY342 | + | 166.50 | 129.52 | 55.97 | 4534 | ! | M90644 Buchnera aphidicola DNA |
| gb_pr4:AC0937845 | + | 166.50 | 129.29 | 57.37 | 4655 | ! | X03831 P.falciparum (CAMP stra |
| gb_in4:PEGP195 | + | 166.50 | 127.74 | 70.30 | 5541 | ! | AF111783 Homo sapiens myosin I |
| gb_pr5:AF111783 | + | 166.50 | 127.01 | 77.19 | 6016 | ! | AF111783 Homo sapiens myosin I |
| gb_pl3:SCU03673 | + | 166.50 | 126.86 | 78.78 | 6125 | ! | AC03673 Saccharomyces cerevisia |
| gb_pat2:E26748 | + | 166.50 | 126.83 | 79.07 | 6145 | ! | E26748 Structure and function |
| gb_pat2:AX067150 | + | 166.50 | 125.95 | 88.45 | 6782 | ! | AX067150 Sequence 2 from Paten |
| gb_ba1:AE007151 | + | 166.50 | 124.57 | 105.68 | 7932 | ! | AX067151 Sequence 3 from Paten |
| gb_in3:CELSPC3A | - | 166.50 | 123.04 | 128.54 | 9424 | ! | M85149 Caenorhabditis elegans |
| gb_ba1:AE000770 | - | 166.50 | 119.60 | 199.90 | 13900 | ! | AE000770 Aquifex aeolicus sec |
| gb_pl3:SC9813 | - | 166.50 | 119.02 | 215.09 | 14825 | ! | Z47814 S.cerevisiae chromosom |
| gb_pl4:VSC88082 | - | 166.50 | 113.35 | 445.23 | 28124 | ! | U10399 Saccharomyces cerevisi |
| gb_pl3:SC354CHV7 | - | 166.50 | 111.30 | 579.23 | 35452 | ! | X95966 S.cerevisiae 35.4kb DN |
| gb_hg15:AC027282 | + | 166.50 | 109.62 | 718.21 | 42839 | ! | AC027282 Homo sapiens chromo |
| gb_pl3:CELF36H12 | + | 166.50 | 109.41 | 738.59 | 43907 | ! | AF078790 Caenorhabditis eleg |
| gb_pl3:SCD8035 | + | 166.50 | 105.40 | 1.2e+03 | 69023 | ! | U33050 Saccharomyces cerevisi |
| gb_pl3:FL003 | + | 166.50 | 104.28 | 1.4e+03 | 78341 | ! | AC06550 Arabidopsis thaliana |
| gb_pr3:AF135798 | + | 166.50 | 101.83 | 2.0e+03 | 103209 | ! | ALJ35798 Human DNA sequence |
| gb_hg1:AC006832 | + | 166.50 | 100.99 | 2.2e+03 | 115522 | ! | AC006832 Homo sapiens chromo |
| gb_hg21:AF1365325 | + | 166.50 | 98.57 | 3.0e+03 | 149125 | ! | AC069401 Homo sapiens chromo |
| gb_hg7:AC021000 | - | 166.50 | 97.75 | 3.3e+03 | 163567 | ! | AL365325 Homo sapiens chromo |
| gb_pl3:CN501DFO | + | 166.00 | 96.35 | 3.9e+03 | 191575 | ! | AC021000 Homo sapiens chromo |
| gb_pr5:AF17292 | + | 166.00 | 142.34 | 10.81 | 1020 | ! | AL16784 Botrytis cinerea stra |
| gb_ba2:AF154023 | + | 166.00 | 132.67 | 28.92 | 2425 | ! | AF17292 Homo sapiens genethon |
| gb_pr4:AF019085 | + | 166.00 | 132.63 | 37.55 | 3052 | ! | AF154023 Streptococcus pneumo |
| gb_pr5:AF127481 | + | 166.00 | 132.48 | 38.31 | 3106 | ! | AF019085 Homo sapiens bromod |
| gb_pr1:AB040942 | + | 166.00 | 128.28 | 65.67 | 4991 | ! | AF127481 Homo sapiens non-ocod |
| gb_r01:MNFMFG | + | 166.00 | 127.77 | 70.05 | 5283 | ! | AB040942 Homo sapiens mRNA for |
| en_ov:GGD547 | + | 166.00 | 127.46 | 72.89 | 5471 | ! | X05640 Mouse NF-M gene for mid |
| gb_ov:AC059399 | + | 166.00 | 126.71 | 80.22 | 5982 | ! | D89547 Chicken mRNA for vitell |
| gb_ov:AF212149 | + | 166.00 | 126.66 | 80.75 | 5987 | ! | AB059399 Bos taurus MyHC-2x m |
| gb_ov:AF212149 | + | 166.00 | 126.65 | 80.91 | 5997 | ! | AF212149 Oryctolagus cuniculu |
| gb_ov:AF212149 | + | 166.00 | 126.59 | 81.53 | 6038 | ! | AF212148 Oryctolagus cuniculu |

| | | | | | | | | | | | | | |
|-------------------|---|--------|--------|---------|--------|-----------------------------------|-------------------|---|--------|--------|---------|--------|-------------------------------|
| gb_hlg6:AC020202 | - | 172.00 | 104.94 | 1.3e+03 | 118591 | ! AC020202 Drosophila melanoga | gb_pr10:HUMAF4Y | + | 170.00 | 125.83 | 89.82 | 9390 | ! LI3773 Human AF-4 mRNA, com |
| gb_in1:AC007837 | - | 172.00 | 102.47 | 1.4e+03 | 123647 | ! AC007837 Drosophila melanoga | gb_pat2:143747 | + | 170.00 | 125.83 | 89.83 | 9391 | ! I43747 Sequence 25 from pat |
| gb_in1:AC009218 | - | 172.00 | 102.47 | 1.8e+03 | 126753 | ! AC009218 Drosophila melanoga | gb_in3:DME245406 | + | 170.00 | 113.66 | 428.03 | 93107 | ! A245406 Drosophila melano |
| gb_hlg11:AC026903 | - | 172.00 | 102.34 | 1.8e+03 | 159179 | ! AC026903 Homo sapiens chromo | gb_hlg5:AC017329 | + | 170.00 | 105.24 | 1.3e+03 | 95961 | ! AC017329 Drosophila melano |
| gb_in2:AC009551 | - | 172.00 | 99.39 | 2.7e+03 | 22079 | ! AC009551 Homo sapiens chromo | gb_pr3:AC001493 | + | 170.00 | 100.34 | 1.6e+03 | 118869 | ! AC001493 Homo sapiens PAC |
| gb_hlg19:AC003791 | - | 172.00 | 97.97 | 3.2e+03 | 260686 | ! AC003791 Drosophila melanoga | gb_pr1:AC011456 | + | 170.00 | 100.92 | 2.2e+03 | 156357 | ! AC011456 Homo sapiens chr |
| gb_ov:HSMA011327 | + | 171.50 | 134.70 | 28.65 | 3926 | ! AL133161 Homo sapiens mRNA; cl | gb_rn1:AC0078930 | + | 170.00 | 97.65 | 3.3e+03 | 226168 | ! AC0078930 Mus musculus 10 |
| gb_ov:XLA8211130 | + | 171.50 | 130.70 | 48.13 | 6198 | ! AJ251130 Xenopus laevis mRNA f | gb_rn1:AC0078930 | + | 170.00 | 97.65 | 3.3e+03 | 226168 | ! AC0078930 Mus musculus 10 |
| gb_ov:AC022655 | + | 171.50 | 128.65 | 62.63 | 7814 | ! AF022655 Homo sapiens cep250 c | gb_rn2:RATNEM | + | 169.50 | 136.43 | 23.09 | 2717 | ! M18628 Rat NF-M middle mole |
| gb_in2:AR093136 | + | 171.50 | 127.82 | 69.60 | 8575 | ! AF093136 Toxoplasma gondii DNA | gb_pat2:AF269940 | + | 169.50 | 132.93 | 31.81 | 3602 | ! AF269940 Staphylococcus epi |
| gb_ov:GGI19187 | + | 171.50 | 122.45 | 138.55 | 15717 | ! Y19187 Gallus gallus mRNA for | gb_in2:AF183401 | + | 169.50 | 132.94 | 37.51 | 4165 | ! AF183401 Caenorhabditis ele |
| gb_ov:SPBC336 | + | 171.50 | 114.29 | 394.92 | 39512 | ! AL121815 S.pombe chromosome 1 | gb_pat1:AS92450 | + | 169.50 | 131.16 | 45.33 | 4920 | ! A92450 Sequence 1 from pate |
| gb_ov:CELU11C4 | + | 171.50 | 113.54 | 434.52 | 42978 | ! U64854 Caenorhabditis elegans | gb_pr10:HUMHNF2B | + | 169.50 | 130.67 | 48.27 | 5200 | ! D26156 Human mRNA for trans |
| gb_ov:AC000547 | + | 171.50 | 104.21 | 1.4e+03 | 123288 | ! AC000547 Homo sapiens genom | gb_pr10:HSU2175 | + | 169.50 | 130.59 | 48.77 | 5247 | ! U29175 Human transcrip |
| gb_ov:AC017582 | + | 171.50 | 103.16 | 1.6e+03 | 133685 | ! AC017582 Drosophila melanoga | gb_ov:AC014886 | + | 169.50 | 130.11 | 51.92 | 5544 | ! AB014886 Emericella nidulan |
| gb_ov:AC019140 | + | 171.50 | 101.21 | 2.1e+03 | 172868 | ! AC019140 Homo sapiens clone | gb_ov:AC0126834 | + | 169.50 | 129.59 | 55.44 | 5874 | ! AF126834 Mus musculus 6rip |
| gb_ov:AC020796 | + | 171.50 | 98.52 | 3.0e+03 | 234125 | ! AC020796 Mus musculus clone | gb_ov:PFANT195 | + | 169.50 | 129.53 | 59.15 | 5917 | ! X02919 Plasmodium falciparu |
| gb_in1:AB003536 | + | 171.50 | 95.18 | 4.6e+03 | 341319 | ! AE003536 Drosophila melanoga | gb_ov:AC020233 | + | 169.50 | 128.99 | 59.95 | 6292 | ! AC020233 Bos taurus mRNA f |
| gb_ov:GLSR2 | + | 171.00 | 132.40 | 38.69 | 4892 | ! X79815 G.lambalia SR2 gene, 671 | gb_in2:PRCP195A | + | 169.50 | 127.99 | 60.35 | 7038 | ! X15063 Plasmodium falciparu |
| gb_ov:CHKVTC | + | 171.00 | 130.90 | 46.87 | 5791 | ! M18060 Chicken vitellogenin ge | gb_in2:AF298215 | + | 169.50 | 127.85 | 69.35 | 7153 | ! AF298215 Drosophila melanog |
| gb_ov:AB002322 | + | 171.00 | 130.75 | 47.82 | 5984 | ! AB002322 Homo sapiens mRNA for | gb_ov:HS85921 | + | 169.50 | 124.27 | 109.70 | 10709 | ! AL049784 Novel human gene |
| gb_ov:NC0243769 | + | 171.00 | 130.62 | 48.63 | 5982 | ! AJ243769 Notothenia coriiceps | gb_ov:AF020713 | + | 169.50 | 101.86 | 1.9e+03 | 134416 | ! AF020713 Bacteriophage SP |
| gb_ov:CAROP2 | + | 171.00 | 130.17 | 51.48 | 6290 | ! Y10377 C.albicans TOP2 gene, 5 | gb_ov:AF021506 | + | 169.50 | 100.23 | 2.4e+03 | 161573 | ! AF021506 Human herpesviru |
| gb_ov:AF201422 | + | 171.00 | 128.28 | 65.64 | 7789 | ! AF201422 Homo sapiens splicing | gb_in1:AC009183 | + | 169.50 | 98.70 | 2.9e+03 | 192055 | ! AC009183 Drosophila melan |
| gb_pat1:AG1387 | + | 171.00 | 128.02 | 67.87 | 8022 | ! A61387 Sequence 3 from Patent | gb_hlg5:AC023833 | + | 169.50 | 98.25 | 3.1e+03 | 202083 | ! AC023833 Mus musculus chr |
| gb_ov:SPRAD3GEN | + | 171.00 | 128.02 | 67.87 | 8022 | ! Y09076 S.pombe RAD3 gene, 4/19 | gb_ov:BSU00012 | + | 169.50 | 97.75 | 3.3e+03 | 213680 | ! Z99115 Bacillus subtilis |
| gb_ov:SPU76307 | + | 171.00 | 128.02 | 67.87 | 8022 | ! U76307 Schizosaccharomyces pom | gb_hlg16:AC074307 | + | 169.50 | 97.48 | 3.4e+03 | 220469 | ! AC074307 Mus musculus chr |
| gb_ov:AB016092 | + | 171.00 | 126.97 | 77.61 | 9027 | ! AB016092 Homo sapiens mRNA for | gb_in2:AE003685 | + | 169.50 | 97.30 | 3.5e+03 | 225038 | ! AE003685 Drosophila melan |
| gb_ov:AE000600 | + | 171.00 | 124.10 | 112.24 | 12489 | ! AE000600 Helicobacter pylori 1 | gb_ov:AF130343 | + | 169.50 | 94.97 | 4.7e+03 | 292721 | ! AF130343 Homo sapiens chr |
| gb_ov:U39701 | + | 171.00 | 122.36 | 140.21 | 15191 | ! U39701 Mycoplasma genitalium | gb_rn1:AC013809 | + | 169.00 | 132.70 | 37.22 | 3956 | ! AF103809 Mus musculus p-3 |
| gb_ov:SPBC216 | + | 171.00 | 117.84 | 250.45 | 25311 | ! AL049558 S.pombe chromosome 1 | gb_in2:AF009782 | + | 169.00 | 128.17 | 66.52 | 6595 | ! U09782 Argopecten irradians |
| gb_ov:AC004493 | + | 171.00 | 113.01 | 465.34 | 43661 | ! AC004493 Homo sapiens chromo | gb_in2:AF220353 | + | 169.00 | 127.01 | 77.27 | 7524 | ! AF220353 Drosophila melanog |
| gb_hlg21:AL359955 | + | 171.00 | 101.35 | 2.1e+03 | 162782 | ! AL359955 Homo sapiens chromo | gb_hlg5:AC017387 | + | 169.00 | 124.79 | 102.74 | 9668 | ! AC017387 Drosophila melanog |
| gb_ov:AF000719 | + | 171.00 | 99.68 | 2.6e+03 | 196424 | ! AP000719 Homo sapiens genom | gb_hlg6:AC017670 | + | 169.00 | 113.21 | 453.40 | 35708 | ! AC017670 Drosophila melanog |
| gb_hlg18:AC087566 | + | 171.00 | 98.21 | 3.1e+03 | 231912 | ! AE007566 Mus musculus clone | gb_hlg6:AC017670 | + | 169.00 | 113.21 | 453.40 | 35708 | ! AC017670 Drosophila melanog |
| gb_hlg22:AL445993 | + | 171.00 | 97.97 | 3.2e+03 | 238351 | ! AL445993 Homo sapiens chromo | gb_hlg24:AP003268 | + | 169.00 | 100.16 | 2.4e+03 | 129290 | ! AC020761 Homo sapiens chr |
| gb_in2:AC003824 | + | 171.00 | 97.58 | 3.4e+03 | 249001 | ! AE003824 Drosophila melanoga | gb_hlg12:AC034201 | + | 169.00 | 97.49 | 3.4e+03 | 210617 | ! AC034201 Homo sapiens chr |
| gb_hlg7:AC020870 | + | 171.00 | 97.32 | 3.5e+03 | 256373 | ! AC020870 Mus musculus clone | gb_in1:AE003485 | + | 168.50 | 94.28 | 5.1e+03 | 302350 | ! AE003485 Drosophila melan |
| gb_rn1:BC003741 | + | 170.50 | 138.89 | 16.83 | 22649 | ! BC003741 Mus musculus. Similar | gb_rn2:RATN0P140B | + | 168.50 | 134.09 | 31.15 | 3235 | ! M94288 Rattus norvegicus nu |
| gb_ov:DMHMCARA | + | 170.50 | 135.26 | 26.81 | 3388 | ! D10667 Homo sapiens mRNA for s | gb_hlg6:AC017858 | + | 168.50 | 131.28 | 44.65 | 4441 | ! AC017858 Drosophila melanog |
| gb_in4:PFPL90 | + | 170.50 | 131.33 | 44.41 | 5282 | ! X03371 P.falciparum gp190 (MS | gb_rn1:MM028789 | + | 168.50 | 129.90 | 53.31 | 5191 | ! U28789 Mus musculus p3 ass |
| gb_ov:SPAC59F8 | + | 170.50 | 114.06 | 40.73 | 37093 | ! Z69728 S.pombe chromosome 1 C | gb_ov:AC025262 | + | 168.50 | 128.92 | 61.26 | 5866 | ! U28789 Mus musculus p3 ass |
| gb_ov:AC005937 | + | 170.50 | 111.90 | 536.41 | 47323 | ! AC005937 Homo sapiens clone t | gb_in2:CEC2507 | + | 168.50 | 111.30 | 579.13 | 42361 | ! Z81039 Caenorhabditis eleg |
| gb_hlg4:AC013847 | + | 170.50 | 111.50 | 564.42 | 49491 | ! AC013847 Drosophila melanoga | gb_ov:AC005750 | + | 168.50 | 104.86 | 1.3e+03 | 87684 | ! AC005750 Drosophila melan |
| gb_hlg1:AC005979 | + | 170.50 | 108.55 | 824.51 | 69084 | ! AC005979 Drosophila melanoga | gb_in1:AC005750 | + | 168.50 | 103.47 | 1.6e+03 | 102516 | ! AC005750 Drosophila melan |
| gb_hlg6:AC018038 | + | 170.50 | 106.81 | 1.0e+03 | 84011 | ! AC018038 Drosophila melanoga | gb_in1:AC005750 | + | 168.50 | 103.47 | 1.6e+03 | 102516 | ! AC005750 Drosophila melan |
| gb_ov:AC079041 | + | 170.50 | 103.70 | 1.5e+03 | 119420 | ! AC079041 Arabidopsis thalian | gb_ov:AC005750 | + | 168.50 | 95.88 | 4.2e+03 | 241429 | ! AE003608 Drosophila melan |
| gb_hlg6:AC020329 | + | 170.50 | 101.39 | 2.1e+03 | 154895 | ! AC020329 Drosophila melanoga | gb_ov:AC005750 | + | 168.50 | 93.63 | 5.6e+03 | 311321 | ! AE003608 Drosophila melan |
| gb_ov:AC084730 | + | 170.50 | 100.73 | 2.2e+03 | 166929 | ! AC084730 Papio hamadryas cl | gb_ov:AC005750 | + | 168.00 | 139.42 | 15.72 | 1695 | ! AF067128 Streptococcus pneu |
| gb_in1:AC009316 | + | 170.50 | 100.61 | 2.3e+03 | 169271 | ! AC009316 Drosophila melanoga | gb_ov:AC005750 | + | 168.00 | 139.16 | 16.26 | 1746 | ! AJ002055 Streptococcus pneu |
| gb_in1:AC008304 | + | 170.50 | 99.97 | 2.5e+03 | 181955 | ! AC008304 Drosophila melanoga | gb_ov:AC005750 | + | 168.00 | 136.05 | 24.23 | 2485 | ! AF019904 Streptococcus pneu |
| gb_hlg7:AC020858 | + | 170.50 | 99.76 | 2.5e+03 | 186222 | ! AC020858 Mus musculus clone | gb_ov:AC005750 | + | 168.00 | 135.22 | 26.96 | 2725 | ! AF068646 Streptococcus pneu |
| gb_ov:AB023048 | + | 170.50 | 99.46 | 2.6e+03 | 192650 | ! AB023048 Homo sapiens genom | gb_in1:CEU19615 | + | 168.00 | 134.96 | 27.88 | 2806 | ! U19615 Caenorhabditis eleg |
| gb_hlg1:AB023048 | + | 170.50 | 99.31 | 2.7e+03 | 195906 | ! AC084071 Mus musculus clone | gb_ov:PSV154012 | + | 168.00 | 130.16 | 30.88 | 3070 | ! AF154012 Streptococcus pneu |
| gb_hlg7:AF000511 | + | 170.50 | 99.13 | 2.8e+03 | 200000 | ! AP000511 Homo sapiens genom | gb_ov:PSV154012 | + | 168.00 | 130.16 | 30.88 | 3070 | ! AF154012 Streptococcus pneu |
| gb_in1:AC008237 | + | 170.50 | 99.12 | 2.8e+03 | 200216 | ! AC008237 Drosophila melanoga | gb_ov:AC008237 | + | 168.00 | 129.53 | 55.88 | 5175 | ! M97501 Human cytoplasmic li |
| gb_hlg22:AL599737 | + | 170.50 | 99.00 | 2.8e+03 | 202939 | ! AL599737 Mus musculus chromo | gb_ov:AC008237 | + | 168.00 | 129.19 | 58.41 | 5380 | ! AC008237 Homo sapiens RNA |
| gb_hlg18:AC087227 | + | 170.50 | 98.88 | 2.8e+03 | 205691 | ! AC087227 Mus musculus chromo | gb_in2:AF056936 | + | 168.00 | 129.12 | 58.90 | 5420 | ! AF056936 Plasmodium falcipa |
| gb_in2:AE0033748 | + | 170.50 | 96.92 | 3.7e+03 | 256608 | ! AE0033748 Drosophila melanoga | gb_ov:AC003460 | + | 168.00 | 127.27 | 74.73 | 6683 | ! U24070 Rattus norvegicus Mu |
| gb_in1:AE003438 | + | 170.50 | 95.60 | 4.4e+03 | 279537 | ! AE003438 Drosophila melanoga | gb_ov:AC003460 | + | 168.00 | 126.62 | 81.24 | 7193 | ! M22286 D.melanogaster beta |
| gb_in2:AF162774 | + | 170.50 | 95.55 | 4.4e+03 | 299507 | ! AE003438 Drosophila melanoga | gb_ov:AC003460 | + | 168.00 | 122.83 | 131.98 | 9443 | ! U35637 Human nebulin mRNA |
| gb_rn2:RATN0P140A | + | 170.00 | 135.89 | 24.73 | 3018 | ! AF162774 Drosophila melanogast | gb_in1:CEU12469 | + | 168.00 | 121.70 | 152.64 | 12530 | ! AE001417 Plasmodium falcip |
| gb_ov:RNNFMR | + | 170.00 | 135.74 | 25.22 | 3070 | ! Z12152 R.norvegicus mRNA for n | gb_ov:RNNFMR | + | 168.00 | 117.17 | 272.70 | 20881 | ! X83957 H.sapiens mRNA for |
| gb_rn10:HUMSPRP | + | 170.00 | 134.30 | 30.30 | 3609 | ! M94287 Rattus norvegicus nucle | gb_hlg4:AC014130 | + | 168.00 | 111.39 | 572.76 | 40323 | ! AC014130 Drosophila melanog |
| gb_ov:CFCS3VS | + | 170.00 | 132.07 | 40.35 | 4643 | ! L25050 Human serine/proline-ri | gb_hlg6:AC020317 | + | 168.00 | 106.54 | 1.1e+03 | 69348 | ! AC020317 Drosophila melanog |
| gb_rn10:HUMTOPII | + | 170.00 | 131.79 | 41.82 | 4792 | ! J04088 Human DNA topoisomerase | gb_hlg12:AC044798 | + | 168.00 | 105.39 | 1.2e+03 | 78927 | ! AC030517 Drosophila melanog |
| gb_rn2:RNNHCB | + | 170.00 | 129.91 | 53.23 | 5932 | ! Y15939 Rat mRNA for beta card | gb_hlg12:AC044798 | + | 168.00 | 102.60 | 1.8e+03 | 108202 | ! AC044798 Homo sapiens chr |
| gb_ov:AB025260 | + | 170.00 | 129.90 | 53.30 | 5932 | ! AB025260 Sus scrofa mRNA for m | gb_hlg3:AC011703 | + | 168.00 | 99.15 | 2.8e+03 | 159777 | ! AC011703 Drosophila melan |
| gb_in2:AINHCHRC | + | 170.00 | 129.79 | 54.08 | 6008 | ! M58018 Homo sapiens beta-myosin | gb_hlg3:AC011703 | + | 168.00 | 99.15 | 2.8e+03 | 159777 | ! AC011703 Drosophila melan |
| gb_ov:AC017582 | + | 170.00 | 128.76 | 61.74 | 6751 | ! X55714 Scallop (A. irradians) m | gb_hlg3:AC017582 | + | 168.00 | 97.04 | 3.6e+03 | 202741 | ! AC017582 Drosophila melan |
| gb_pat2:143748 | + | 170.00 | 125.85 | 89.60 | 9390 | ! I43748 Sequence 27 from patent | gb_hlg19:AF276759 | + | 168.00 | 96.50 | 3.9e+03 | 215478 | ! AF276759 Homo sapiens chr |
| gb_in1:AX015531 | + | 170.00 | 125.83 | 89.82 | 9390 | ! AX015531 Sequence 225 from Pat | gb_in1:AX015531 | + | 168.00 | 93.54 | 5.7e+03 | 300994 | ! AE003506 Drosophila melan |

| | | | | | | | | | | | | | | | |
|---------------------|---|--------|--------|---------|--------|------------------------------|-------------------------|---------------------|---|--------|--------|---------|--------|-----------|----------------------|
| gb_hcg24:CEY52B11 | - | 178.00 | 103.66 | 1.5e+03 | 233941 | 298867 | Caenorhabditis elegans | gb_in3:CEL004D1 | - | 175.00 | 119.89 | 192.52 | 28678 | AF067617 | Caenorhabditis el |
| gb_hcg1:AC006279 | - | 178.00 | 103.22 | 1.6e+03 | 245802 | AC006279 | Plasmodium falciparum | gb_hcg1:AC006710 | - | 175.00 | 105.48 | 1.2e+03 | 145920 | AC006710 | Caenorhabditis e |
| gb_hcg1:AC006702 | - | 178.00 | 101.46 | 2.0e+03 | 298864 | AC006702 | Caenorhabditis elegans | gb_pr2:AC006101 | - | 175.00 | 105.33 | 1.2e+03 | 148278 | AC006101 | cib338 f 24, c |
| gb_pr1:AB046824 | + | 177.50 | 141.58 | 11.92 | 3088 | AB046824 | Homo sapiens mRNA for | gb_in1:AC008218 | + | 175.00 | 103.68 | 1.5e+03 | 178792 | AC008218 | Drosophila melan |
| gb_in3:D63884 | + | 177.50 | 139.69 | 15.18 | 3832 | D63884 | Anthocidaris crassispin | gb_pr2:AC007461 | + | 175.00 | 103.60 | 1.6e+03 | 180385 | AC007461 | Homo sapiens chr |
| gb_pr9:HSN0AMR | + | 177.50 | 134.16 | 30.86 | 7154 | Z11583 | H.sapiens mRNA for Numa | gb_in1:AC007892 | + | 175.00 | 102.55 | 1.8e+03 | 202929 | AC007892 | Drosophila melan |
| gb_pr9:HSN0DMR | + | 177.50 | 129.26 | 57.89 | 12446 | X14298 | Human mRNA for dystrop | gb_in2:AE003770 | + | 175.00 | 101.67 | 2.0e+03 | 224100 | AE003770 | Drosophila melan |
| gb_hcg4:AC014744 | + | 177.50 | 122.25 | 142.21 | 27448 | AC014744 | Drosophila melanogast | gb_hcg24:CEY79H2 | + | 175.00 | 100.72 | 2.3e+03 | 249640 | AE002287 | Caenorhabditis e |
| gb_in1:AC005811 | + | 177.50 | 113.59 | 432.05 | 72987 | AC005811 | Drosophila melanogast | gb_hcg1:AC0079433 | + | 175.00 | 100.61 | 2.3e+03 | 252689 | AC0079433 | Mus musculus chr |
| gb_in2:AC003669 | + | 177.50 | 101.91 | 1.9e+03 | 272605 | AC003669 | Drosophila melanogast | gb_in3:CEY75B8A | + | 175.00 | 99.14 | 2.8e+03 | 298406 | AC003514 | Caenorhabditis e |
| gb_pr7:BC006094 | + | 177.00 | 142.16 | 11.07 | 2775 | BC006094 | Homo sapiens, clone M | gb_hcg24:CEY75B8A | + | 175.00 | 98.07 | 3.2e+03 | 336638 | AC003514 | Caenorhabditis e |
| gb_pr1:AF026255 | + | 177.00 | 141.60 | 11.90 | 2957 | AF026255 | Mus musculus plenty-C | gb_in4:PFMEZSALB | + | 174.50 | 134.43 | 29.81 | 5312 | Z35327 | P.falciparum gp190 (|
| gb_hcg2:AG28541 | + | 177.00 | 138.40 | 17.93 | 4243 | G28541 | human STS SHGC-31621, S | gb_in4:PFMEZSALB | + | 174.50 | 132.62 | 37.62 | 6519 | AR016730 | Sequence 1 from pa |
| gb_pr7:D63875 | + | 177.00 | 134.91 | 17.93 | 4243 | D63875 | Human mRNA for KIAA0155 | em.pat1:AR016730 | + | 174.50 | 132.62 | 37.62 | 6519 | AR016730 | Sequence 1 from pa |
| gb_pr1:AB003476 | + | 177.00 | 134.91 | 28.03 | 6287 | AB003476 | Homo sapiens mRNA for | gb_pr10:H0UMDOCK180 | + | 174.50 | 132.62 | 37.62 | 6519 | AR016730 | Sequence 1 from pa |
| gb_pr4:AF001504 | + | 177.00 | 134.91 | 28.08 | 6297 | AF001504 | Homo sapiens myasthen | gb_in1:AF318573 | + | 174.50 | 129.53 | 55.91 | 9238 | AB035419 | Drosophila melanog |
| gb_pr10:HUMBP4 | + | 177.00 | 134.06 | 31.26 | 6921 | M63618 | Human bullous pemphig | gb_v11:AF318573 | + | 174.50 | 129.53 | 55.91 | 9238 | AB035419 | Drosophila melanog |
| gb_in2:AF221715 | + | 177.00 | 125.69 | 91.46 | 17802 | AF221715 | Drosophila melanogast | gb_hcg20:ALL59162 | + | 174.50 | 103.93 | 1.5e+03 | 156234 | AF318573 | Bovine herpesvir |
| gb_in2:AF184612 | + | 177.00 | 125.38 | 95.14 | 18430 | AF184612 | Drosophila melanogast | gb_hcg20:ALL59162 | + | 174.50 | 103.93 | 1.5e+03 | 156234 | AF318573 | Bovine herpesvir |
| gb_in2:AF188205 | + | 177.00 | 124.91 | 101.11 | 19444 | AF188205 | Drosophila melanogast | gb_hcg25:PFMAL13P3 | + | 174.00 | 98.17 | 3.1e+03 | 318221 | AF159162 | Homo sapiens chr |
| gb_hcg1:AC005646 | + | 177.00 | 111.67 | 552.52 | 86677 | AC005646 | Drosophila melanogast | gb_in1:AE003429 | + | 174.00 | 133.96 | 31.69 | 5261 | AS9333 | Sequence 2 from Pate |
| gb_in1:AC005334 | - | 177.00 | 109.44 | 735.18 | 111448 | AC005334 | Drosophila melanogast | gb_in4:PFAT7010 | + | 174.00 | 133.68 | 32.82 | 5529 | AJ007010 | Plasmodium falcipa |
| gb_hcg9:AC024486 | - | 177.00 | 106.20 | 1.1e+03 | 160903 | AC024486 | Homo sapiens chromo | gb_in4:SCYGR098C | + | 174.00 | 133.32 | 34.40 | 5763 | Z27883 | S.cerevisiae chromos |
| gb_in1:AC008371 | - | 177.00 | 106.10 | 1.1e+03 | 160903 | AC024486 | Homo sapiens chromo | gb_in3:DMU77947 | + | 174.00 | 129.98 | 52.79 | 8401 | U77947 | Drosophila melanog |
| gb_hcg2:AC019340 | - | 177.00 | 105.02 | 1.3e+03 | 183481 | AC019340 | Homo sapiens chromo | gb_hcg4:AC014378 | + | 174.00 | 125.33 | 95.86 | 14201 | AC014378 | Drosophila melanog |
| gb_in1:AE003814 | + | 177.00 | 101.91 | 1.9e+03 | 260139 | AE003814 | Drosophila melanogast | gb_hcg5:AC016449 | + | 174.00 | 125.33 | 95.86 | 14201 | AC014378 | Drosophila melanog |
| gb_in2:AE003590 | + | 177.00 | 100.62 | 2.3e+03 | 307139 | AE003590 | Drosophila melanogast | gb_in1:AE003429 | + | 174.00 | 125.33 | 95.86 | 14201 | AC014378 | Drosophila melanog |
| gb_in2:AF215635 | + | 176.50 | 132.45 | 38.59 | 7967 | AF215635 | Drosophila melanogast | gb_in1:AE003429 | + | 174.00 | 125.33 | 95.86 | 14201 | AC014378 | Drosophila melanog |
| gb_in3:CELMAYNC | + | 176.50 | 131.34 | 44.32 | 9000 | J01050 | Caenorhabditis elegans | gb_in1:AE003429 | + | 174.00 | 125.33 | 95.86 | 14201 | AC014378 | Drosophila melanog |
| gb_hcg20:ALL58197 | + | 176.50 | 115.07 | 357.09 | 56457 | ALL58197 | Homo sapiens chromos | gb_pr2:AC007722 | + | 173.50 | 110.11 | 675.21 | 75703 | AC007722 | Homo sapiens chro |
| gb_hcg24:CEL10666_3 | + | 176.50 | 109.16 | 761.95 | 110000 | Continuation (4 of 7) of CEX | | gb_hcg1:AC0079414 | + | 173.50 | 102.89 | 1.7e+03 | 170916 | AC0079414 | Homo sapiens chr |
| gb_pr2:ATF24M12 | + | 176.50 | 107.72 | 917.32 | 129516 | AL132980 | Arabidopsis thalian | gb_hcg6:AC019152 | + | 173.50 | 102.41 | 1.8e+03 | 180520 | AC019152 | Homo sapiens chr |
| gb_hcg20:AL135364 | + | 176.50 | 106.40 | 1.1e+03 | 129516 | AL135364 | Homo sapiens chromo | gb_hcg18:AC087226 | + | 173.50 | 98.97 | 2.1e+03 | 265985 | AC087226 | Mus musculus chr |
| gb_in1:AC010843 | + | 176.50 | 105.26 | 1.3e+03 | 170913 | AC010843 | Drosophila melanogast | gb_ba2:AC002995 | + | 173.00 | 96.62 | 3.8e+03 | 346897 | AC002995 | Mesohizobium lo |
| gb_hcg2:AC010994 | + | 176.50 | 104.53 | 1.4e+03 | 185573 | AC010994 | Drosophila melanogast | gb_r2:RNDNATPII | + | 173.00 | 132.81 | 32.30 | 4987 | Z19552 | R.norvegicus mRNA fo |
| gb_hcg9:AC023477 | + | 176.50 | 104.02 | 1.5e+03 | 196536 | AC023477 | Homo sapiens chromo | gb_pr5:AF227948 | + | 173.00 | 133.23 | 34.78 | 5323 | AF227948 | Homo sapiens HBV p |
| gb_hcg13:AC055134 | + | 176.50 | 103.66 | 1.5e+03 | 204603 | AC066591 | Homo sapiens chromo | gb_r2:RNDNATPII | + | 173.00 | 132.09 | 40.24 | 6052 | Z46372 | R.norvegicus RNA for |
| gb_hcg20:AL135513 | + | 176.50 | 101.12 | 2.1e+03 | 272677 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 117.62 | 257.55 | 31002 | U39847 | Caenorhabditis eleg |
| gb_in1:AE003492 | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_hcg4:AC014947 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295.10 | 34946 | AC014947 | Drosophila melanog |
| gb_pr10:HUMRGSLU | + | 176.50 | 100.31 | 2.4e+03 | 298839 | AL355134 | Homo sapiens chromo | gb_in3:CEU39847 | + | 173.00 | 116.56 | 295. | | | |

| | | | | | | | | | | | |
|-------------------|----------|--------|---------|--------|------------------------------------|-------------------|----------|--------|---------|--------|--------------------------------|
| gb_ro1:AF013969 | + 185.50 | 143.73 | 9.05 | 4957 | ! AF013969 Mus musculus antigen | gb_ba3:MMU47023 | + 181.00 | 135.63 | 25.58 | 8285 | ! U47023 Methanococcus marip |
| gb_p10:HUMNP220 | - 185.50 | 141.23 | 12.47 | 6571 | ! D83032 Homo sapiens mRNA for n | gb_p14:YCES91 | - 181.00 | 135.19 | 27.04 | 8700 | ! D11088 Saccharomyces cerevi |
| gb_in3:CELF32E10 | - 185.50 | 126.85 | 78.85 | 33318 | ! U41992 Caenorhabditis elegans | gb_sy:AF134159 | - 181.00 | 131.85 | 41.52 | 12689 | ! AF134159 Expression shuttl |
| gb_pat1:AX063459 | + 185.00 | 149.17 | 4.50 | 2565 | ! AX063459 Sequence 4 from Paten | gb_p11:SC29711 | + 181.00 | 124.32 | 109.07 | 29866 | ! X75891 S.cerevisiae 29.7Kb |
| gb_in2:AB042160 | + 185.00 | 147.79 | 5.38 | 2998 | ! AF042160 Plasmodium falciparum | gb_in3:CELUK03C7 | + 181.00 | 124.23 | 110.34 | 29989 | ! U40059 Caenorhabditis eleg |
| gb_in2:AB042160 | + 185.00 | 147.37 | 5.67 | 3142 | ! AB042160 Rattus norvegicus mRN | gb_htg18:AC087689 | + 181.00 | 106.60 | 1.1e+03 | 219259 | ! X15309 H.sapiens NF-H gene, |
| gb_ov:PMU19361 | + 185.00 | 145.78 | 6.96 | 3763 | ! U19361 Petromyscus marinus neur | gb_pr3:HSGNFH4 | + 180.50 | 143.97 | 7.03 | 2542 | ! L06487 Saccharomyces cerevi |
| gb_ov:PMU19361 | + 185.00 | 144.40 | 8.30 | 4393 | ! AF051357 Mus musculus golgin-2 | gb_p14:YSCZ1P1A | + 180.50 | 145.70 | 8.78 | 3091 | ! M69183 Plasmodium falciparu |
| gb_om:SSU75316 | + 185.00 | 141.65 | 11.82 | 5996 | ! U75316 Sus scrofa beta-myosin | gb_in4:PFAMESA | + 180.50 | 139.71 | 15.14 | 4995 | ! L40157 Human endosome-associ |
| gb_ba1:AR006525 | + 185.00 | 137.09 | 21.20 | 10029 | ! AE006525 Streptococcus pyogen | gb_p10:HUMERAP | + 180.50 | 139.66 | 15.26 | 5028 | ! L40157 Human endosome-associ |
| gb_p10:HSU91931 | + 184.50 | 143.84 | 6.90 | 3574 | ! U91931 Homo sapiens AP-3 compl | gb_ov:AF285155 | + 180.50 | 139.43 | 15.71 | 5158 | ! AF285155 Gallus gallus topo |
| gb_pr3:HSP162 | + 184.50 | 143.26 | 9.61 | 4780 | ! X78998 H.sapiens p162 mRNA. 3/ | gb_in1:AB015484 | + 180.50 | 138.05 | 18.75 | 6029 | ! AB015484 Dugesia japonica m |
| gb_ov:AB007445 | + 184.50 | 141.58 | 11.92 | 5781 | ! AB007445 Gallus gallus mRNA f | gb_in2:AF166170 | + 180.50 | 135.94 | 24.56 | 7646 | ! AF166170 Caenorhabditis ele |
| gb_ro1:MUSDYSA | - 184.50 | 133.86 | 32.09 | 13815 | ! M68859 Mouse dystrophin mRNA, | gb_in2:AF166169 | + 180.50 | 135.96 | 25.06 | 7781 | ! AF166169 Caenorhabditis ele |
| gb_pat1:AC005617 | - 184.50 | 130.89 | 46.94 | 19307 | ! AC005617 Arabidopsis thaliana | gb_in2:AF261891 | + 180.50 | 134.96 | 27.86 | 8543 | ! AF261891 Caenorhabditis ele |
| gb_p11:AC005617 | - 184.50 | 125.42 | 94.74 | 35822 | ! AC005617 Arabidopsis thaliana | gb_pr3:AF160252 | + 180.50 | 132.24 | 39.49 | 11611 | ! AF160252 Homo sapiens KIAA |
| gb_p10:NCB13118 | - 184.50 | 121.01 | 166.76 | 58918 | ! AL390189 Neurospora crassa DN | gb_htg6:AC020215 | + 180.50 | 125.47 | 94.14 | 24942 | ! U51031 Saccharomyces cerev |
| gb_htg16:AC079219 | - 184.50 | 111.07 | 596.97 | 181008 | ! AC079219 Mus musculus chromo | gb_htg6:AC020215 | + 180.50 | 118.22 | 238.43 | 56510 | ! AC020215 Drosophila melan |
| gb_htg6:AC019525 | + 184.50 | 109.72 | 709.29 | 210663 | ! AC019525 Drosophila melanoga | gb_pr1:AC000025 | + 180.50 | 110.48 | 643.56 | 135405 | ! AC000025 Homo sapiens Chr |
| gb_pr6:AL3355497 | - 184.50 | 109.52 | 727.43 | 215397 | ! AL335497 Human DNA sequence | gb_pr2:AC005527 | + 180.50 | 109.61 | 719.15 | 149308 | ! AC005527 Homo sapiens 22q |
| gb_in1:AE003481 | + 184.50 | 106.02 | 1.1e+03 | 319850 | ! AE003481 Drosophila melanoga | gb_p11:AC025806 | - 180.50 | 109.26 | 752.70 | 155420 | ! AC025806 Oryza sativa chr |
| gb_p14:SCYL1021W | + 184.00 | 141.86 | 11.50 | 5354 | ! X73126 S.cerevisiae chromosome | gb_in1:AE003551 | + 180.50 | 103.86 | 1.5e+03 | 285978 | ! AE003551 Drosophila melan |
| gb_p13:SCSPAG | + 184.00 | 140.79 | 13.19 | 6041 | ! X53731 S. cerevisiae SPA2 gene | gb_in1:AE003550 | + 180.50 | 103.51 | 1.6e+03 | 297385 | ! AE003550 Drosophila melan |
| gb_p10:HSU72938 | + 184.00 | 136.14 | 23.96 | 10216 | ! U72938 Homo sapiens putative | gb_p13:SCCHRII1 | + 180.50 | 102.99 | 1.7e+03 | 315339 | ! X59720 S.cerevisiae chrom |
| gb_p10:HSU72937 | + 184.00 | 136.04 | 24.26 | 10330 | ! U72937 Homo sapiens putative | gb_pr2:AC005529 | + 180.50 | 102.90 | 1.7e+03 | 318488 | ! AC005529 Homo sapiens chr |
| gb_p10:HSU72936 | + 184.00 | 135.94 | 24.59 | 10452 | ! U72936 Homo sapiens putative | gb_pat2:EL4401 | + 180.00 | 146.37 | 6.45 | 2254 | ! EL4401 cDNA encoding HET-B |
| gb_in2:AP297644 | + 184.00 | 135.48 | 26.06 | 11000 | ! AF297644 Drosophila melanogas | gb_p10:HSU81504 | + 180.00 | 141.40 | 12.20 | 3950 | ! U81504 Homo sapiens beta-3A |
| gb_p13:SC32KBF | + 184.00 | 126.00 | 87.94 | 32088 | ! X97560 S.cerevisiae 32kb DNA | gb_in2:AR247194 | + 180.00 | 139.70 | 15.17 | 4784 | ! AF247194 Drosophila melanog |
| gb_ov:GGNFM10 | + 183.50 | 141.01 | 12.83 | 5638 | ! X17102 Chicken DNA for middle- | gb_om:OCU32574 | + 180.00 | 137.71 | 19.57 | 5988 | ! U32574 Oryctolagus cuniculu |
| gb_in1:AE001424 | + 183.50 | 134.14 | 30.95 | 12240 | ! AE001424 Plasmodium falciparu | gb_p10:HSU53786 | + 180.00 | 137.04 | 21.32 | 6457 | ! U53786 Homo sapiens envopla |
| gb_p13:NCB1109 | + 183.50 | 118.56 | 228.40 | 71085 | ! AL513409 Neurospora crassa DN | gb_pr3:HSN00AMRB | + 180.00 | 136.06 | 24.20 | 7217 | ! U11584 H.sapiens mRNA for N |
| gb_htg14:AC068279 | - 183.50 | 112.77 | 479.57 | 136555 | ! AC068279 Homo sapiens chromo | gb_in3:DMU91980 | + 180.00 | 135.80 | 25.02 | 7431 | ! U19180 Drosophila melanogas |
| gb_om:AB234765 | + 183.00 | 152.43 | 2.96 | 1485 | ! AF234765 Rattus norvegicus sex | gb_pat1:AR031367 | + 180.00 | 134.77 | 28.56 | 8351 | ! AR031367 Sequence 14 from p |
| gb_om:AB234765 | + 183.00 | 140.16 | 14.31 | 5938 | ! AB234765 Sus scrofa MYHC-slow | gb_pat2:168238 | + 180.00 | 134.77 | 28.56 | 8351 | ! 168238 Sequence 14 from pat |
| gb_pat2:AX088804 | + 183.00 | 139.65 | 15.26 | 6284 | ! AX088804 Sequence 756 from Pat | gb_p13:SCU1109 | + 180.00 | 132.86 | 36.48 | 10358 | ! U31109 Saccharomyces cerevi |
| gb_pr3:HSB001013 | + 183.00 | 139.65 | 15.26 | 6284 | ! AX088804 Sequence 756 from Pat | gb_p11:AR162221 | + 180.00 | 107.63 | 928.05 | 178733 | ! U75653 Human zinc finger h |
| gb_in3:AC081376 | - 183.00 | 134.05 | 31.31 | 11829 | ! AE001376 Plasmodium falciparu | gb_st52:ONS07DYV | + 179.50 | 153.25 | 2.67 | 992 | ! AL441104 T7 end of clone B00 |
| gb_in3:CH38K22 | - 183.00 | 127.21 | 75.32 | 25610 | ! AL024499 Caenorhabditis elega | gb_p10:BC004286 | + 179.50 | 140.50 | 13.68 | 4179 | ! BC004286 Homo sapiens' clon |
| gb_htg14:AC073693 | + 183.00 | 107.74 | 913.93 | 230372 | ! AC073693 Mus musculus clone | gb_pr5:AF214114 | + 179.50 | 137.59 | 19.87 | 5804 | ! AF214114 Homo sapiens' retin |
| gb_p10:HSU91930 | + 182.50 | 141.75 | 11.66 | 4744 | ! U91930 Homo sapiens AP-3 compl | gb_om:AB059400 | + 179.50 | 137.40 | 20.36 | 5930 | ! AB059400 Bos taurus MYHC-s1 |
| gb_p13:SCU17580 | + 182.50 | 139.85 | 14.88 | 5877 | ! U19580 Saccharomyces cerevisia | gb_pat1:AR018129 | + 179.50 | 136.86 | 21.84 | 6306 | ! AR018129 Sequence 3 from pa |
| gb_p10:HUMYFOHP | + 182.50 | 139.64 | 15.28 | 6016 | ! M36769 Homo sapiens perinatal | gb_pat2:173445 | + 179.50 | 136.86 | 21.84 | 6306 | ! 173445 Sequence 3 from pate |
| gb_pat1:AR019266 | + 182.50 | 139.23 | 16.12 | 6306 | ! AR019266 Homo sapiens chromo | gb_pat2:185731 | + 179.50 | 136.86 | 21.84 | 6306 | ! 185731 Sequence 3 from pate |
| gb_htg4:AC012625 | + 182.50 | 111.93 | 534.24 | 137363 | ! AC012625 Homo sapiens chromo | gb_p10:AC000035 | + 179.50 | 120.85 | 170.24 | 38429 | ! AC000035 Homo sapiens Chro |
| gb_p10:AC018634 | - 182.50 | 109.82 | 699.99 | 174241 | ! AC018634 Human Chromosome 7 | gb_htg16:AC074305 | + 179.50 | 105.77 | 1.2e+03 | 216677 | ! AC074305 Mus musculus chr |
| gb_ro1:DC3850 | + 182.00 | 147.95 | 5.27 | 2254 | ! D63850 Mus musculus mRNA for h | gb_htg16:AC074305 | + 179.50 | 103.49 | 1.6e+03 | 272677 | ! AL355134 Homo sapiens chr |
| gb_in2:AF145640 | + 182.00 | 147.92 | 5.28 | 2260 | ! AF145640 Drosophila melanogast | gb_in1:AE003543 | - 179.50 | 103.12 | 1.7e+03 | 284398 | ! AE003543 Drosophila melan |
| gb_p10:AF340028 | + 182.00 | 142.37 | 10.77 | 4230 | ! AF340028 Mus musculus Rab6-int | gb_ro1:MMU63648 | + 179.00 | 140.22 | 14.19 | 4126 | ! U63648 Mus musculus p160 my |
| gb_p10:AB015617 | + 182.00 | 140.99 | 12.86 | 4943 | ! AB015617 Homo sapiens ELKS mRN | gb_p13:SC2958X | + 179.00 | 120.40 | 180.30 | 38661 | ! Z49259 S.cerevisiae chromo |
| gb_pr3:HSRRP129 | + 182.00 | 140.36 | 13.94 | 5307 | ! Y11251 H.sapiens mRNA for nove | gb_p13:NC99H12 | + 179.00 | 113.40 | 442.31 | 85165 | ! AL451018 Neurospora crassa |
| gb_in2:AF134172 | + 182.00 | 139.48 | 15.61 | 5863 | ! AF134172 Pecten maximus myosin | gb_htg1:AC009091 | + 179.00 | 106.24 | 1.1e+03 | 191140 | ! AC009091 Homo sapiens chr |
| gb_ov:AF004397 | + 182.00 | 138.07 | 18.69 | 6872 | ! AF004397 Gallus gallus chromo- | gb_htg4:AC014439 | - 178.50 | 131.38 | 44.09 | 10705 | ! AC002967 Rattus norvegicus |
| gb_p11:AC000015 | + 182.00 | 117.94 | 247.32 | 66704 | ! AC000015 Homo sapiens chromo- | gb_p13:AT28119 | - 178.50 | 125.56 | 93.04 | 20656 | ! AC014439 Drosophila melano |
| gb_htg24:AP003235 | - 182.00 | 110.47 | 644.54 | 154975 | ! AP003235 Oryza sativa chromo | gb_p13:AT28119 | - 178.50 | 110.68 | 627.21 | 10766 | ! AC035709 Arabidopsis thal |
| gb_htg18:AC084878 | - 182.00 | 109.81 | 701.02 | 176946 | ! AC084878 Homo sapiens chromo | gb_htg12:AC044873 | - 178.50 | 106.01 | 1.1e+03 | 187554 | ! AC044873 Homo sapiens chr |
| gb_in1:AC008216 | - 182.00 | 109.17 | 761.57 | 169886 | ! AC008216 Drosophila melanoga | gb_htg1:AC009501 | - 178.50 | 104.65 | 1.4e+03 | 218687 | ! AC009501 Mus musculus cio |
| gb_htg24:AP003286 | - 182.00 | 109.08 | 769.84 | 181200 | ! AC003286 Oryza sativa chromo | gb_htg1:AC009501 | - 178.50 | 103.55 | 15.47 | 4074 | ! M94389 Squid neurofilament |
| gb_in2:AC003759 | - 182.00 | 107.11 | 991.86 | 226471 | ! AC003759 Drosophila melanoga | gb_p11:AB018346 | + 178.00 | 136.67 | 22.39 | 5639 | ! AB018346 Homo sapiens mRNA |
| gb_htg2:AC001072 | - 182.00 | 106.75 | 1.0e+03 | 235711 | ! AC010172 Homo sapiens chromo | gb_p11:AS8932 | + 178.00 | 135.89 | 24.71 | 6152 | ! AS8932 Sequence 1 from Pate |
| gb_pr4:AF048977 | + 181.50 | 143.17 | 9.72 | 3698 | ! AF048977 Homo sapiens Ser/Arg- | gb_in2:AR324831 | + 178.00 | 132.44 | 38.52 | 9091 | ! AF324831 Plasmodium falcipa |
| gb_p10:AF043102 | + 181.50 | 142.47 | 10.64 | 4002 | ! AF043102 Pneumocystis carini | gb_in2:AR324831 | + 178.00 | 131.43 | 43.81 | 10181 | ! AF277386 Drosophila melano |
| gb_ba1:AF143911 | + 181.50 | 139.28 | 16.02 | 5737 | ! AF143911 Mycoplasma pneumoniae | gb_in3:HSANK840 | + 178.00 | 129.60 | 55.40 | 12518 | ! Z26634 Homo sapiens mRNA f |
| gb_pr3:HSWYOSIN | + 181.50 | 138.86 | 16.89 | 6010 | ! Z38133 H.sapiens mRNA for myos | gb_in3:CELUK09G12 | - 178.00 | 121.29 | 160.82 | 33179 | ! AF047663 Caenorhabditis el |
| gb_ba1:SCWYOIG | + 181.50 | 138.72 | 17.20 | 6108 | ! X53947 Yeast MYO1 gene for myc | gb_in3:CFP2809 | - 178.00 | 120.83 | 170.72 | 33706 | ! Z81518 Caenorhabditis eleg |
| gb_ba1:AF143912 | - 181.50 | 137.44 | 20.28 | 7080 | ! AF143912 Mycoplasma pneumoniae | gb_in3:CELC42D8 | + 178.00 | 118.71 | 223.96 | 42000 | ! U56366 Caenorhabditis eleg |
| gb_p10:HUMDYS | + 181.50 | 131.40 | 43.99 | 13957 | ! M18533 Homo sapiens dystroph | gb_htg24:CEV53HL1 | - 178.00 | 110.35 | 654.59 | 118001 | ! Continuation (2 of 4) of |
| gb_in3:CELY54E10A | + 181.50 | 113.81 | 419.83 | 101640 | ! AC024810 Caenorhabditis eleg | gb_htg4:AC013398 | - 178.00 | 106.58 | 1.1e+03 | 168216 | ! AC013398 Homo sapiens chr |
| gb_htg9:AC023572 | - 181.50 | 111.19 | 587.58 | 136630 | ! AC023572 Homo sapiens chromo | gb_htg11:AC026495 | + 178.00 | 105.67 | 1.2e+03 | 186506 | ! AC026495 Homo sapiens chr |
| gb_p11:AB002530 | + 181.00 | 142.72 | 10.30 | 3760 | ! AB002530 Neurospora crassa mus | gb_htg2:AC010321 | - 178.00 | 104.53 | 1.4e+03 | 212074 | ! AC010321 Homo sapiens chr |
| gb_pr4:AF091711 | + 181.00 | 139.00 | 16.59 | 5660 | ! AF091711 Homo sapiens splice v | | | | | | |
| gb_pat1:AR002603 | + 181.00 | 137.63 | 19.77 | 6605 | ! AR002603 Sequence 4 from paten | | | | | | |
| gb_p10:HSU81607 | + 181.00 | 137.63 | 19.77 | 6605 | ! HSU81607 Homo sapiens gravin mRN | | | | | | |
| gb_p10:HSU81607 | + 181.00 | 136.23 | 23.68 | 6608 | ! Z36005 S.cerevisiae chromosome | | | | | | |

| | | | | | | | | | | | |
|--------------------|--------|--------|--------|--------|---------------------------------|-------------------|--------|--------|--------|--------|-------------------------------|
| gb.p13:SCNUM1 | 202.00 | 150.67 | 3.72 | 9851 | X61236 S.cerevisiae NUM1 gene, | gb.ov:AF027728 | 192.00 | 142.99 | 9.94 | 9610 | AF027728 Xenopus laevis kin |
| gb.p13:SC8338 | 201.00 | 137.52 | 20.07 | 43468 | U50046 S.cerevisiae chromosome | gb.in2:CBRG42F08 | 192.00 | 129.32 | 47.45 | 44999 | AC084596 Caenorhabditis br |
| gb.p12:SC831906 | 201.50 | 153.35 | 2.64 | 6965 | Z13906 Homo sapiens golgin-245 | gb.in2:AC087748 | 192.00 | 125.44 | 94.52 | 69740 | AC087748 Homo sapiens chr |
| gb.pat2:135495 | 201.50 | 150.02 | 4.04 | 10136 | I35495 Sequence 2 from patent | gb.hg4:AC012857 | 192.00 | 117.64 | 266.79 | 168062 | AC012857 Drosophila melan |
| gb.p10:10HUMGCP372 | 201.50 | 149.88 | 4.11 | 10295 | D25542 Human mRNA for golgi 2 | gb.pr3:AC010723 | 192.00 | 117.31 | 267.85 | 174420 | AC010723 Homo sapiens BAC |
| gb.p14:SPRC27 | 201.00 | 148.95 | 4.63 | 10943 | L1353866 S.pombe chromosome 1 | gb.ro1:AF194970 | 191.50 | 143.81 | 8.96 | 8385 | AF194970 Mus musculus BAC1 |
| gb.in4:PFARHPR | 200.50 | 152.52 | 2.93 | 6991 | L27838 Plasmodium yoelii rhop | gb.in1:AE003558 | 191.50 | 112.93 | 470.22 | 273744 | AE003558 Drosophila melan |
| gb.in4:PFARHPR | 200.50 | 152.80 | 2.83 | 6484 | AF335500 Physarum polycephalum | gb.in1:AE003558 | 191.00 | 149.02 | 4.59 | 4455 | AF340183 Homo sapiens BAC |
| gb.in2:AF232343 | 200.00 | 150.67 | 3.71 | 8241 | AF323443 Plasmodium yoelii yo | gb.pr9:HSMP67 | 191.00 | 141.59 | 11.93 | 10302 | X69086 H.sapiens mRNA for |
| gb.p14:SPRC162 | 200.00 | 137.14 | 21.07 | 37960 | AF323860 S.pombe chromosome 1 | gb.pat1:AC03607 | 191.00 | 141.57 | 11.93 | 10320 | AF3607 Sequence 9 from Pat |
| gb.p11:AB016871 | 200.00 | 130.63 | 48.53 | 79109 | BC0018871 Arabidopsis thaliana | gb.v12:U97553 | 191.00 | 119.88 | 192.78 | 119450 | U97553 Murine herpesvirus |
| gb.pr7:BC001883 | 199.50 | 160.97 | 0.9911 | 2464 | BC001883 Homo sapiens, nucleol | gb.v11:AF105037 | 191.00 | 119.87 | 192.96 | 119550 | AF105037 Murid herpesviru |
| gb.ro1:BC003993 | 199.50 | 158.97 | 1.28 | 3089 | BC003993 Mus musculus, clone M | gb.hg12:AC037485 | 191.00 | 117.66 | 256.37 | 135317 | AC037485 Homo sapiens chr |
| gb.in2:AF196347 | 199.50 | 156.32 | 1.80 | 4164 | AF196347 Plasmodium falciparum | gb.in1:AC007853 | 191.00 | 117.13 | 274.29 | 182921 | AC007853 Drosophila melan |
| gb.in2:AF196347 | 199.50 | 118.35 | 234.65 | 302855 | AF003420 Drosophila melanoga | gb.in1:AC008206 | 191.00 | 116.19 | 309.38 | 181132 | AC008206 Drosophila melan |
| gb.ov:AF286475 | 199.00 | 160.90 | 1.00 | 2376 | AF286475 Takifugu rubripes ret | gb.in2:AE003750 | 191.00 | 114.17 | 400.64 | 227398 | AE003750 Neisseria mening |
| gb.in2:AF134186 | 199.00 | 155.59 | 1.98 | 4326 | AF134186 Caenorhabditis elegans | gb.b3:NNAG622491 | 191.00 | 110.98 | 603.88 | 326301 | AF162757 Neisseria mening |
| gb.ro2:RATGCP60 | 199.00 | 148.02 | 5.22 | 10170 | D25543 Rattus norvegicus mRNA | gb.on:AF165917 | 190.50 | 150.39 | 3.85 | 3649 | AF165917 Canis familiaris t |
| gb.in3:CELTG031 | 199.00 | 136.99 | 21.48 | 35318 | AF067614 Caenorhabditis elegans | gb.in3:DR0M0XNA | 190.50 | 145.50 | 7.21 | 6338 | M35012 D.melanogaster gon-m |
| gb.in1:AC008140 | 199.00 | 124.73 | 103.53 | 140973 | AC008140 Drosophila melanoga | gb.pr10:HUMSEQX | 190.50 | 141.99 | 11.31 | 9416 | L06237 Human microtubule-as |
| gb.in1:AC009219 | 199.00 | 123.66 | 118.70 | 159007 | AC009219 Drosophila melanoga | gb.hg6:AC020480 | 190.50 | 138.81 | 57.01 | 13487 | AC020480 Drosophila melan |
| gb.in3:CEU10412 | 198.50 | 154.66 | 2.23 | 4597 | U10412 Caenorhabditis elegans | gb.in2:CEC25A1 | 190.50 | 129.04 | 19.49 | 40597 | 281038 Caenorhabditis eleg |
| gb.in1:AC008140 | 198.00 | 152.46 | 2.95 | 5894 | U00624 Acanthamoeba castellan | gb.in3:DME271740 | 190.50 | 124.17 | 111.20 | 70398 | AJ271740 Drosophila melan |
| gb.ov:GGU62026 | 198.00 | 152.13 | 3.08 | 5850 | U62026 Gallus gallus cardiac m | gb.hg6:AC017682 | 190.50 | 123.57 | 120.07 | 75313 | AC017682 Drosophila melan |
| gb.pr9:HSMPAC | 197.50 | 146.72 | 6.16 | 10300 | X75304 H.sapiens giantin mRNA | gb.hg2:CEX3986_1 | 190.50 | 120.21 | 184.66 | 110000 | Continuation (2 of 4) of |
| gb.pat2:181218 | 197.00 | 147.73 | 5.41 | 8789 | I81218 Sequence 5 from patent | gb.p11:AC009323 | 190.50 | 117.97 | 246.32 | 141748 | AC009323 Arabidopsis thal |
| gb.pat2:HSU30872 | 197.00 | 146.40 | 6.42 | 10211 | U30872 Human mitotin mRNA, co | gb.hg1:AC012609 | 190.50 | 117.56 | 259.45 | 148380 | AC012609 Homo sapiens chr |
| gb.in3:CEU12F3 | 197.00 | 135.11 | 27.35 | 36555 | U80022 Caenorhabditis elegans | gb.hg7:AC020738 | 190.50 | 116.65 | 291.75 | 164520 | AC020738 Homo sapiens chr |
| gb.in3:CEU34201 | 196.50 | 165.74 | 0.5375 | 2033 | U34201 Oryctolagus cuniculus c | gb.hg7:AC024601 | 190.50 | 116.26 | 306.58 | 171857 | AC024601 Homo sapiens chr |
| gb.bal:AF053765 | 196.00 | 147.62 | 5.49 | 8142 | AF053765 Bacillus megaterium A | gb.in3:CEY398B | 190.50 | 114.61 | 378.80 | 207022 | AL132896 Caenorhabditis e |
| gb.pr10:HSU19769 | 196.00 | 145.71 | 7.91 | 10096 | U19769 Human CENP-F kinetoch | gb.pr2:AC008174 | 190.50 | 120.35 | 181.58 | 103663 | AC008174 Homo sapiens BAC |
| gb.in3:DMC43E4 | 196.00 | 133.57 | 33.28 | 37949 | AL031128 Drosophila melanoga | gb.pr2:HSU18985 | 189.50 | 151.49 | 3.34 | 2947 | U18985 Human triadin mRNA, |
| gb.hg18:ACN087189 | 196.00 | 119.97 | 190.43 | 184494 | AC087189 Homo sapiens chromo | gb.ov:GCLAU5A | 189.50 | 150.30 | 3.89 | 3371 | X67778 G.gallus mRNA for cl |
| gb.sts1:ACN06XES | 195.50 | 165.74 | 0.5375 | 1007 | AL196139 T3 end of clone XAX0A | gb.pat1:AC048216 | 189.50 | 147.05 | 5.91 | 4868 | AF048216 Sequence 12 from p |
| gb.bal:AB002503 | 195.50 | 144.20 | 8.52 | 11455 | AC02503 Neisseria meningitid | gb.in2:AF036278 | 189.50 | 143.73 | 9.05 | 7079 | AF036278 Drosophila melanog |
| gb.in3:CEU06A7 | 195.50 | 136.53 | 22.79 | 72423 | Z78066 Caenorhabditis elegans | gb.in2:AF036278 | 189.50 | 139.97 | 14.65 | 10816 | Y18278 Drosophila melanog |
| gb.p12:AF020307 | 195.50 | 127.75 | 70.28 | 73391 | AP020307 Arabidopsis thaliana | gb.hg2:AC009160 | 189.50 | 114.54 | 382.17 | 130856 | AC009160 Homo sapiens chr |
| gb.p13:SC43817 | 195.50 | 126.44 | 83.06 | 75317 | Z38059 S.cerevisiae chromosome | gb.b4:ECR751TRA | 189.00 | 146.53 | 6.32 | 4936 | X59794 E.coli broad-host-ra |
| gb.in1:AB018113 | 195.50 | 126.44 | 83.06 | 85020 | AL08113 Arabidopsis thaliana | gb.pr4:AF083037 | 189.00 | 138.33 | 16.22 | 11318 | AF083037 Homo sapiens A-k1 |
| gb.pat1:AX044033 | 195.00 | 113.91 | 414.63 | 349980 | AX044033 Sequence 112 from F | gb.pr8:HSAL31693 | 189.00 | 138.32 | 18.10 | 12465 | AJ131693 Homo sapiens mRNA |
| gb.on:BTCYCN1 | 195.00 | 159.24 | 1.24 | 2007 | Z46789 B.taurus mRNA for cyclic | gb.p14:SPAC1F5 | 189.00 | 127.94 | 68.56 | 40242 | Z68136 S. pombe chromosome |
| gb.pr5:AK026473 | 195.00 | 154.86 | 2.17 | 3289 | AK026473 Homo sapiens cDNA: FL | gb.b2:BAU67194 | 189.00 | 125.44 | 94.44 | 53339 | U67194 Enterobacter aeroge |
| gb.hg1:AC006280 | 195.00 | 120.26 | 183.62 | 163443 | AC006280 Plasmodium falcipar | gb.hg6:AC019755 | 189.00 | 122.01 | 146.70 | 78598 | AC019755 Drosophila melan |
| gb.in1:AC006281 | 195.00 | 118.49 | 230.36 | 195551 | AC006281 Plasmodium falcipar | gb.hg2:AC010344 | 189.00 | 116.12 | 312.31 | 152829 | AC010344 Homo sapiens chr |
| gb.on:RADNAA | 194.50 | 157.22 | 1.60 | 2410 | M94315 Rabbit neurofilament-H | gb.p12:AF233527 | 188.50 | 148.18 | 5.11 | 3916 | AF233527 Arabidopsis thaliana |
| gb.on:RADNAA | 194.50 | 151.00 | 3.56 | 4655 | L10065 Oryctolagus cuniculus c | gb.p12:AF233527 | 188.50 | 148.18 | 5.11 | 3916 | AF233527 Arabidopsis thaliana |
| gb.b3:RFU47030 | 194.00 | 150.18 | 3.96 | 5103 | U47030 Rhizobium tropici putat | gb.on:AF176816 | 188.50 | 145.79 | 6.95 | 5134 | AF176816 Bos taurus KIA037 |
| gb.on:AB009387 | 194.00 | 149.16 | 4.51 | 5724 | AB009387 Sus scrofa mRNA for | gb.hg9:AC024036 | 188.50 | 114.20 | 399.37 | 181485 | AC024036 Homo sapiens chr |
| gb.p1:AB002371 | 194.00 | 148.80 | 4.72 | 5967 | AB002371 Human mRNA for KIAA03 | gb.ro2:RNU83590 | 188.00 | 144.12 | 8.61 | 5929 | AB025261 Sus scrofa mRNA fo |
| gb.hg1:AC008184 | 194.00 | 120.12 | 186.92 | 151872 | AC008184 Drosophila melanoga | gb.ov:GMYGSM | 188.00 | 143.22 | 9.65 | 6556 | X05346 Chicken mRNA for giz |
| gb.hg12:AC036204 | 194.00 | 119.56 | 200.78 | 161742 | AC036204 Homo sapiens chromo | gb.in2:AF233442 | 188.00 | 141.18 | 12.54 | 8256 | AF233442 Plasmodium yoelii |
| gb.pr3:AC009487 | 194.00 | 118.44 | 231.98 | 183666 | AC009487 Homo sapiens BAC cl | gb.ro1:MMY12229 | 188.00 | 138.56 | 17.55 | 11096 | Y12229 M.musculus mRNA for |
| gb.in2:AE003659 | 194.00 | 115.33 | 345.67 | 260897 | AE003659 Drosophila melanoga | gb.p13:CE23A5 | 188.00 | 118.27 | 237.09 | 109694 | AC011713 Arabidopsis thal |
| gb.in2:AF217802 | 193.50 | 150.92 | 3.60 | 4490 | AF217802 Drosophila melanoga | gb.pr9:HSMP210 | 187.50 | 155.70 | 1.95 | 1533 | M94315 Rabbit neurofilament |
| gb.v12:HSV3PRGEN | 193.50 | 130.77 | 47.69 | 43658 | M86409 Herpesvirus salmiriti th | gb.on:RSGMAP210 | 187.50 | 142.97 | 9.97 | 6452 | Y12490 Homo sapiens mRNA fo |
| gb.hg4:AC015193 | 193.50 | 130.54 | 49.13 | 44815 | AC015193 Drosophila melanoga | gb.in3:DDIMVHC | 187.50 | 142.66 | 10.37 | 6681 | M14628 D.discoidium myosin |
| gb.hg9:AC013932 | 193.50 | 122.31 | 141.42 | 112930 | AC013932 Drosophila melanoga | gb.bal:MPUS9896 | 187.50 | 137.41 | 20.34 | 12083 | U59896 Mycoplasma pneumonia |
| gb.hg1:AC008284 | 193.50 | 120.03 | 189.17 | 146797 | AC008284 Drosophila melanoga | gb.bal:AE000051 | 187.50 | 134.79 | 28.48 | 16249 | AE000051 Mycoplasma pneumo |
| gb.in2:AE003752 | 193.50 | 116.22 | 308.07 | 225478 | AE003752 Drosophila melanoga | gb.ro1:MUSNLF5A | 187.00 | 149.38 | 4.06 | 2797 | M24496 Mouse neurofilament |
| gb.sts1:CN50797T | 193.00 | 164.33 | 0.6441 | 945 | AL434943 T7 end of clone BB0A0 | gb.hg5:AC016135 | 186.50 | 131.36 | 44.24 | 22988 | U46675 Caenorhabditis eleg |
| gb.pat1:A92451 | 193.00 | 149.68 | 4.22 | 4940 | A92451 Sequence 2 from Patent | gb.pat1:AC054896 | 186.50 | 141.48 | 12.07 | 6981 | AC054896 Sequence 11 from P |
| gb.sy:CEV131294 | 193.00 | 149.68 | 4.22 | 4940 | AJ131294 Cloning vector pBSK+ | gb.pr10:HUMBPAG1B | 186.50 | 139.55 | 15.46 | 8684 | L11690 Human bullous 210 kd |
| gb.pr4:AFM030234 | 193.00 | 149.06 | 4.57 | 5298 | AF030234 Homo sapiens splicing | gb.in1:AE003091 | 186.50 | 139.50 | 15.96 | 8930 | M69225 Human bullous pemphi |
| gb.b3:MHMP34GN | 193.00 | 146.85 | 6.06 | 6797 | X95601 M.hominis lmp3 and lmp4 | gb.hg1:AC006368 | 186.50 | 115.53 | 336.64 | 130659 | AC006368 Homo sapiens chr |
| gb.hg18:AF184623 | 193.00 | 143.36 | 9.48 | 10077 | AF184623 Plasmodium vivax ret | gb.hg5:AC0016135 | 186.50 | 112.68 | 485.23 | 180249 | AC0016135 Homo sapiens chr |
| gb.in3:CELT8973 | 192.50 | 118.45 | 231.45 | 167671 | AC090081 Homo sapiens chromo | gb.pr7:BC000755 | 186.00 | 150.91 | 3.60 | 2305 | BC000755 Homo sapiens, simi |
| gb.in1:AC084463 | 192.50 | 130.30 | 50.67 | 42126 | AC006834 Caenorhabditis elegans | gb.ro1:AF245512 | 186.00 | 148.08 | 8.65 | 4380 | AF245512 Rattus norvegicus |
| gb.in1:XLNOPP180 | 192.00 | 124.01 | 113.48 | 83644 | AC084463 Caenorhabditis briggs | gb.in2:AF240482 | 186.00 | 142.60 | 10.46 | 5888 | AF230482 Hydra vulgaris tig |
| gb.ro1:AB004305 | 192.00 | 152.12 | 3.09 | 3432 | X89527 X.laevis mRNA for XNPP | gb.in4:PFARPCPA | 186.00 | 139.00 | 16.60 | 8841 | M89097 Plasmodium vivax ret |
| gb.ov:CCU53862 | 192.00 | 147.27 | 4.44 | 4730 | AB004305 Mus musculus mRNA for | gb.in1:AC087076 | 186.00 | 119.10 | 212.90 | 83495 | AC087076 Caenorhabditis br |
| gb.ro1:AF139185 | 192.00 | 147.23 | 5.77 | 5957 | U53862 Coturnix coturnix slow | gb.ro1:MUSNPH | 185.50 | 145.72 | 7.01 | 3959 | M35131 Mouse neurofilament |
| | 192.00 | 145.83 | 6.91 | 6975 | AF139185 Rattus norvegicus myd | gb.ro1:MMNPHGE | 185.50 | 145.72 | 7.01 | 3960 | Z31012 M.musculus (Swiss-We |

| Sequence | Strd Orig | Zscore | EScore | Len | Documentation |
|--------------------|-----------|---------|----------|--------|----------------------------------|
| gb_pr10.HS06HZ31 | + 7155.00 | 5841.68 | 5.2e-306 | 74444 | AL137201 Novel human gene map |
| gb_pr8.HS095825 | + 7101.00 | 5602.15 | 8.3e-304 | 5253 | U95825 Human androgen-induced |
| gb_pr1.AB023193 | + 6994.00 | 5517.93 | 4.1e-299 | 5110 | AB023196 Homo sapiens MRNA for |
| gb_pr10.HS050533 | + 2953.50 | 2337.44 | 5.8e-122 | 1852 | U50533 Human BRCA2 region, mRNA |
| gb_pr1.AB014548 | + 2704.50 | 2331.78 | 1.7e-110 | 5177 | AB014548 Homo sapiens MRNA for |
| gb_in1.AC007475 | + 1729.50 | 2331.02 | 6.7e-66 | 173613 | AC007475 Drosophila melanogaster |
| gb_in1.AC007474 | + 1729.50 | 1330.09 | 7.5e-66 | 192763 | AC007474 Drosophila melanogaster |
| gb_in2.AC003823 | + 1729.50 | 1327.35 | 1.1e-65 | 262731 | AC003823 Drosophila melanogaster |
| gb_htg6.AC020286 | + 1708.50 | 1313.43 | 6.4e-65 | 194634 | AC020286 Drosophila melanogaster |
| gb_htg5.AC026889 | + 925.50 | 733.56 | 9.8e-33 | 2079 | AK026889 Homo sapiens CDNA: FLJ |
| gb_pr8.HS49J310 | + 925.50 | 698.45 | 1.1e-30 | 137246 | U284572 Human DNA sequence fr |
| gb_pr5.AL1338820 | + 925.50 | 696.63 | 1.4e-30 | 168487 | AL1338820 Human DNA sequence fr |
| gb_pr4.SPAC | + 833.50 | 649.04 | 6.5e-28 | 9982 | AL441624 S.pombe chromosome I |
| gb_htg14.AC068224 | + 819.50 | 622.97 | 1.8e-26 | 54398 | AC068224 Homo sapiens chromos |
| gb_htg92.AL512630 | + 769.00 | 566.85 | 2.4e-23 | 340969 | AL512630 Mus musculus chromo |
| gb_htg5.AC016449 | + 768.00 | 579.81 | 4.6e-24 | 72157 | AC016449 Homo sapiens clone R |
| gb_pr5.AC021757 | + 706.50 | 562.14 | 4.5e-23 | 2212 | AK021757 Homo sapiens cDNA FLJ |
| gb_pr4.SMAJ9934 | + 697.50 | 545.61 | 3.7e-22 | 6411 | AJ009934 Sordaria macrospora S |
| gb_pl3.EMB19 | + 635.00 | 498.68 | 1.5e-19 | 4885 | L03200 Emericella nidulans bic |
| gb_htg14.AC068224 | + 554.50 | 413.79 | 8.2e-15 | 54398 | AC068224 Homo sapiens chromos |
| gb_htg14.AC068352 | + 466.00 | 335.21 | 2.0e-10 | 145450 | AC068352 Homo sapiens chromo |
| gb_htg1.AL1358892 | + 401.50 | 281.68 | 1.9e-07 | 195380 | AL135892 Mus musculus chromo |
| gb_htr8.HS267P19 | + 390.50 | 277.80 | 3.1e-07 | 113704 | U75889 Human DNA sequence fr |
| gb_htg70.AL1353724 | + 390.50 | 274.57 | 4.7e-07 | 163642 | AL1353724 Homo sapiens chromo |
| gb_pl3.SCS582X | + 377.00 | 276.70 | 3.5e-07 | 38661 | U29259 S.cerevisiae chromosome |
| gb_pr5.AC023592 | + 375.50 | 304.37 | 1.0e-08 | 1489 | AK023592 Homo sapiens cDNA FLJ |
| gb_htg8.AC023006 | + 353.50 | 245.41 | 2.0e-05 | 162775 | AC023006 Homo sapiens clone |
| gb_htg14.AC068352 | + 331.00 | 228.65 | 0.0002 | 145450 | AC068352 Homo sapiens chromo |
| gb_htg18.AC090081 | + 326.00 | 223.44 | 0.0003 | 167671 | AC090081 Homo sapiens chromo |
| gb_htg8.AC028682 | + 314.00 | 214.33 | 0.0011 | 160962 | AC028682 Homo sapiens chromo |
| gb_pl1.AFO49529 | + 309.50 | 252.62 | 7.8e-06 | 1431 | AFO49529 Schizosaccharomyces f |
| gb_htg40.AL1356724 | + 277.00 | 184.98 | 0.0456 | 76368 | AL1353724 Homo sapiens chromo |
| gb_pr1.AB019502 | + 248.50 | 190.02 | 0.0239 | 136042 | AB019602 Homo sapiens IDN3-B m |
| gb_pr1.AB019494 | + 234.50 | 184.68 | 0.0474 | 7821 | AB019494 Homo sapiens IDN3 mRN |
| gb_pr4.YSCMLP | + 236.50 | 180.96 | 0.0764 | 6978 | L01992 Saccharomyces cerevisia |
| gb_pl4.SCVKRO95W | + 235.50 | 178.90 | 0.0994 | 8050 | Z28320 S.cerevisiae chromosome |
| gb_pl3.SCDNACH21 | + 235.50 | 172.92 | 0.2142 | 15820 | X73541 S.cerevisiae DNA of ch |
| gb_in3.CEUB038X | + 234.00 | 163.75 | 0.6940 | 38946 | U23168 Caenorhabditis elegans |
| gb_ov2.ALFU12916 | + 232.50 | 174.79 | 0.1684 | 9800 | AF312916 Plasmodium falciparum |
| gb_ov2.XINJ32916 | + 232.50 | 176.95 | 0.1338 | 7321 | Y07624 X.laevis mRNA for NuMA |

THIS PAGE BLANK (USPTO)

shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 267P19 is at 1 in this sequence. The true right end of clone 26423 is at 24539.

The true left end of clone 49J10 is at 113606.

267P19 is from the human PAC library.

FEATURES

```

source
1. .113704
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-13"
/clone="XX-267P19"
repeat_region
1. .235
/partial
/note="Alu repeat: matches 252. .1 of consensus"
repeat_region
370. .662
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
4849. .5140
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
5367. .5660
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
6080. .6187
/partial
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
8133. .8401
/partial
/note="2 copies of 54 mer 87 % conserved"
repeat_region
8623. .8748
/partial
/note="MER7A element fragment"
repeat_region
9617. .9711
/partial
/note="L1 element fragment"
repeat_region
9724. .9876
/partial
/note="Alu repeat: matches 305. .210 of consensus"
repeat_region
10567. .10594
/partial
/note="Alu repeat: matches 158. .1 of consensus"
repeat_region
10616. .10900
/partial
/note="14 copies of 2 mer 93 % conserved"
repeat_region
10960. .10985
/partial
/note="Alu repeat: matches 300. .1 of consensus"
repeat_region
11912. .12100
/partial
/note="13 copies of 2 mer 92 % conserved"
repeat_region
17050. .17359
/partial
/note="Alu repeat: matches 268. .65 of consensus"
repeat_region
17639. .17888
/partial
/note="Alu repeat: matches 1. .306 of consensus"
repeat_region
19151. .19980
/partial
/note="L1 element fragment"
repeat_region
19981. .20264
/partial
/note="Alu repeat: matches 34. .306 of consensus"
repeat_region
20298. .21190
/partial
/note="L1 element fragment"
repeat_region
21199. .21488
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
21538. .22226
/partial
/note="L1 element fragment"
repeat_region
25696. .25989
/partial
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
26902. .27167

```

```

repeat_region
27850. .27930
/note="2 copies of 133 mer 93 % conserved"
repeat_region
28002. .28043
/note="L1 element fragment"
repeat_region
28323. .28780
/note="L1 element fragment"
repeat_region
29176. .29449
/note="Alu repeat: matches 1. .306 of consensus"
repeat_region
29904. .30192
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
30354. .30663
/note="L1 element fragment"
repeat_region
30683. .30972
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
30977. .31204
/note="L1 element fragment"
repeat_region
34819. .35121
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
37059. .37282
/note="MLT2B2 element fragment"
repeat_region
37453. .37506
/note="MLT2B2 element fragment"
repeat_region
38414. .38706
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
39516. .39563
/note="24 copies of 2 mer 81 % conserved"
repeat_region
41244. .41535
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
43994. .44242
/partial
/note="Alu repeat: matches 44. .308 of consensus"
repeat_region
44702. .44889
/note="match: 217151 DNA segment containing (CA) repeat"
repeat_region
44702. .44889
/note="16 copies of CA 100% conserved; Polymorphic by comparison with 217151"
repeat_region
45935. .46229
/note="2 copies of 94 mer 85 % conserved"
repeat_region
46910. .46988
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
47081. .47119
/note="L1 element fragment"
repeat_region
47168. .47431
/note="L1 element fragment"
repeat_region
47615. .47902
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
47922. .48213
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
49436. .49717
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region
50248. .50533
/partial
/note="Alu repeat: matches 308. .1 of consensus"
repeat_region
51097. .51380
/note="Alu repeat: matches 1. .304 of consensus"
repeat_region
62318. .62543
/partial
/note="Alu repeat: matches 302. .66 of consensus"
repeat_region
63358. .63613
/partial
/note="Alu repeat: matches 306. .34 of consensus"
repeat_region
66191. .66236
/note="23 copies of 2 mer 80 % conserved"
repeat_region
68771. .69062
/partial

```


| | | | | | |
|-----------|----------|------------|-----|--|--|
| RESULT | 14 | | | | |
| AC016449 | | | | | |
| LOCUS | 72157 bp | DNA | HTG | | |
| AC016449 | | | | | |
| ACCESSION | | | | | |
| VERSION | 2 | GI:9104517 | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |
| TITLE | | | | | |
| JOURNAL | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |

| | | | |
|-------|--------|------------------|-----------|
| 758 | 857: | contig of 757 bp | in length |
| 858 | 1635: | contig of 100 bp | |
| 1636 | 1735: | gap of 100 bp | |
| 1736 | 2490: | contig of 755 bp | in length |
| 2491 | 2590: | gap of 100 bp | |
| 2591 | 3343: | contig of 753 bp | in length |
| 3344 | 3443: | gap of 100 bp | |
| 3444 | 4224: | contig of 781 bp | in length |
| 4225 | 4324: | gap of 100 bp | |
| 4325 | 5088: | contig of 764 bp | in length |
| 5089 | 5188: | gap of 100 bp | |
| 5189 | 5968: | contig of 780 bp | in length |
| 5969 | 6068: | gap of 100 bp | |
| 6069 | 6844: | contig of 776 bp | in length |
| 6845 | 6944: | gap of 100 bp | |
| 6945 | 7701: | contig of 757 bp | in length |
| 7702 | 7801: | gap of 100 bp | |
| 7802 | 8573: | contig of 772 bp | in length |
| 8574 | 8673: | gap of 100 bp | |
| 8674 | 9430: | contig of 757 bp | in length |
| 9431 | 9530: | gap of 100 bp | |
| 9531 | 10295: | contig of 765 bp | in length |
| 10296 | 10395: | gap of 100 bp | |
| 10396 | 11139: | contig of 744 bp | in length |
| 11140 | 11239: | gap of 100 bp | |
| 11240 | 11988: | contig of 749 bp | in length |
| 11989 | 12088: | gap of 100 bp | |
| 12089 | 12874: | contig of 786 bp | in length |
| 12875 | 12974: | gap of 100 bp | |
| 12975 | 13750: | contig of 776 bp | in length |
| 13751 | 13950: | gap of 100 bp | |
| 13951 | 14638: | contig of 788 bp | in length |
| 14639 | 14738: | gap of 100 bp | |
| 14739 | 15520: | contig of 782 bp | in length |
| 15521 | 15620: | gap of 100 bp | |
| 15621 | 16381: | contig of 761 bp | in length |
| 16382 | 16481: | gap of 100 bp | |
| 16482 | 17248: | contig of 767 bp | in length |
| 17249 | 17348: | gap of 100 bp | |
| 17349 | 18123: | contig of 775 bp | in length |

| | | | |
|--|----------------|---|--------|
| misc_feature | 301501..303502 | /note="assembly_fragment:01554" | |
| misc_feature | 303603..306492 | /note="assembly_fragment:02456" | |
| misc_feature | 306593..309852 | /note="assembly_fragment:03413" | |
| misc_feature | 309953..312398 | /note="assembly_fragment:03967" | |
| misc_feature | 312499..315866 | /note="assembly_fragment:04102" | |
| misc_feature | 315967..318753 | /note="assembly_fragment:04448" | |
| misc_feature | 318854..323190 | /note="assembly_fragment:04616" | |
| misc_feature | 323291..329923 | /note="assembly_fragment:04963" | |
| misc_feature | 330024..332405 | /note="assembly_fragment:06188" | |
| misc_feature | 332506..335631 | /note="assembly_fragment:06207" | |
| misc_feature | 335732..338304 | /note="assembly_fragment:06594" | |
| misc_feature | 338405..340969 | /note="assembly_fragment:06804" | |
| BASE COUNT 100857 a 71110 c 68176 g 96314 t 4512 others | | | ORIGIN |
| Query Match 7.6%; Score 399; DB 81; Length 340969; | | | |
| Best Local Similarity 73.3%; Pred. No. 1.2e-64; | | | |
| Matches 729; Conservative 0; Mismatches 215; Indels 51; Gaps 15; | | | |
| QY | 4270 | icaaattatttgcacaaagttcctaattgttaacatacatattgctgtattaaattc | 4329 |
| DB | 43949 | TAAAGATGTTTAAATAATGATGGCTGGTACTTAGGCTAAATTCGGATGTTTCACATG | 43890 |
| QY | 4330 | catattatggcccccattacactaggtacgcgcggaagtgcgttaaaaggaaagcgcatg | 4389 |
| DB | 43889 | GTCATCTGTTGACTTTTCTCTTCAGGTTACGACGCGGAAGCTCCAAACGAGAGACGATG | 43830 |
| QY | 4390 | acaaatgttaataaactctctctgaaagcttttggaaaaatcttttttttttt | 4449 |
| DB | 43829 | AACACGTC-----ATTAACATCTATTGTGAAGCTTTGGGGGAAAAACATTTTGTGTC | 43775 |
| QY | 4450 | tttttttgcgaagctgagctgaataaagccttgatgcacaaatggagctgta | 4509 |
| DB | 43774 | T-----AAATATGAGGCTAAATAAGCTTGGATGGACACCAACCAAGCTGCCAA | 43726 |
| QY | 4510 | agagtgacagtgtagccttactttgtgaccccatcacatttgggtcacatgctttagc | 4569 |
| DB | 43725 | AGCGTGACAGTTGGAC-----TTGCGATGACCATGGTGTGGTGCACAGGTATCAGC | 43672 |
| QY | 4570 | catacaca-----tggtaacattgactatggagcttctgaaagtgttaatgtgcgatgcta | 4626 |
| DB | 43671 | CATCCACACAGTGATAACACTGACTCCAGAGTCTTGTGAAGTGCTGCTGTGATGCTG | 43612 |
| QY | 4627 | tgtagacataaagaagaactgtgaatatcttttctttttttttttttttttttt | 4686 |
| DB | 43611 | TGTAGACACAACTAAAGAGACACATGTAATAANTCCCACTCCCCCA--TTTCTGACTT | 43554 |
| QY | 4687 | ctgaagtgctgtatagctttttatctcgcgcttttaaaactgacagtacccgactgttatt | 4746 |
| DB | 43553 | CTAAAGTGCTGTATAGCTTTTATCTCGGCTTTAACTACACAGGACCCAGCTGCTACC | 43494 |
| QY | 4747 | ggatctattgatttgaagaatttcttaggatagattcttaagca-----gtaatcttctca | 4802 |
| DB | 43493 | GGATCTGTTTATTGAAAGCATTTGTAGGTGGATCTTAAAGCAGTACGTAGTCTGTCA | 43434 |
| QY | 4803 | gtgtttgatttatttctgaatttactgtgaaaaaaatttttttcaacaattg | 4862 |
| DB | 43433 | GTGTTGTATTGTAATCTCTGCAATTTTACTGTG-----AAATTTGTTTTCACAAATG | 43378 |
| QY | 4863 | gtgtcattttcttgatgcactatttgttggag---agttaaatggctctctccctttgt | 4919 |

| | | | |
|---|-------|---|-------|
| Db | 43377 | GTGTCAATTTTCTCAATGTCACCTGTTTCTGGAGTTAAATGGTTTCTTCTGCCCCCTCTTT | 43318 |
| QY | 4920 | gtatcttaacctagtgttttaactctctggcacccttaactctcaagagtgctaaattgtctg | 4979 |
| Db | 43317 | GTAGCTCATCTA-TGTTTACTCTCAGGCGACTGTT-GTCTTCTGAGGTGCTAAAGCATCTG | 43260 |
| QY | 4980 | ccattacaccagaagtgctctgataggagacacacatgcacaaattgtgaaatagtc | 5039 |
| Db | 43259 | CCAGCACACTAGAA-CATGGCTCTGCTGGGCGACGACATGG---ATTGTGAGATAGTCC | 43204 |
| QY | 5040 | tgaagttcttggattactcttacacctcagttattgttcccagaattttctggccttt | 5099 |
| Db | 43203 | TGAGCTTATCTGATGACTTCACACCTCAGTATTG---GTCCCAGAAATCTCTGGCCTTT | 43148 |
| QY | 5100 | catggcaatgaaaatttttaagaagaagatttaagatttttaatttaagaagtggtt | 5159 |
| Db | 43147 | CATGGCAACGAAAATTTTAAGAA-AGAGATTTAAATATTTTAAAGAGTGTGTT | 43089 |
| QY | 5160 | ataaataatgtactgaattcttccatcttccatcttccatctc-cttcagtttttattaat | 5217 |
| Db | 43088 | ATAAATAATGTAATAATTTCTTTATTCATTCATTCATCTTTTTCAGTTTATATTAAT | 43029 |
| QY | 5218 | ctactgtatcaataaaaattctgtaatttgaatgag | 5252 |
| Db | 43028 | CTACTGTATCAATAAAATTTCTGTAATTTGAATGG | 42994 |
| RESULT 13 | | | |
| AK021757 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |
| FEATURES | | | |
| source | | | |
| Location/Qualifiers | | | |
| 1. 2212 | | | |
| /organism="Homo sapiens" | | | |
| /db_xref="taxon:9606" | | | |
| /clone="HEMBA1005019" | | | |
| /clone_lib="HEMBA1" | | | |
| /dev_stage="embryo, 10 weeks" | | | |
| /tissue_type="whole embryo, mainly head" | | | |
| /note="cloning vector: pME18SFL3" | | | |
| BASE COUNT 751 a 385 c 430 g 646 t | | | |
| AK021757 2212 bp mRNA PRI 29-SEP-2000 | | | |
| Homo sapiens cDNA FLJ11695 fis, clone HEMBA1005019, highly similar | | | |
| to Homo sapiens mRNA for KIAA0648 protein. | | | |
| AK021757 | | | |
| AK021757.1 GI:10433003 | | | |
| Oligo capping; fis (full insert sequence). | | | |
| Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to | | | |
| mRNA, clone_lib:HEMBA1 clone:HEMBA1005019. | | | |
| Homo sapiens | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| 1 (sites) | | | |
| Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., | | | |
| Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., | | | |
| Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., | | | |
| Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., | | | |
| Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A. | | | |
| NEDO human cDNA sequencing project | | | |
| Unpublished (2000) | | | |
| 2 (bases 1 to 2212) | | | |
| Isogai,T. and Otsuki,T. | | | |
| Direct Submission | | | |
| Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao | | | |
| Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, | | | |
| Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, | | | |
| Tel:81-438-52-3951, Fax:81-438-52-3952) | | | |
| NEDO human cDNA sequencing project supported by Ministry of | | | |
| International Trade and Industry of Japan; cDNA full insert | | | |
| sequencing: Research Association for Biotechnology; cDNA library | | | |
| construction, 5'- & 3'-end one pass sequencing and clone selection: | | | |
| Helix Research Institute (supported by Japan Key Technology Center | | | |
| etc.) and Department of Virology, Institute of Medical Science, | | | |
| University of Tokyo. | | | |

DEFINITION Mus musculus chromosome 5 clone RP21-583E8, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.

ACCESSION AL512630
VERSION AL512630.2 GI:13443466
KEYWORDS HTG: HTGS-PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Sims,S.
REFERENCE 1 (bases 1 to 340969)
AUTHORS Direct Submission
TITLE Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 15A, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
COMMENT On Mar 24, 2001 this sequence version replaced gi:12193249.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: dm583E8
----- Summary Statistics
Assembly program: XMAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 323281 bases at least Q40
Consensus quality: 329435 bases at least Q30
Consensus quality: 333234 bases at least Q20
Insert size: 336469; sum-of-contigs
Insert size: 189900; 1.3% error; agarose-fp
Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality
coverage: 11.68x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1..340969
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP21-583E8"
/clone_lib="RPCI-21"
1..10854
/note="assembly_fragment:06116
fragment_chain:1"
10955..16884
/note="assembly_fragment:05022
fragment_chain:1"
16985..22344
/note="assembly_fragment:04089
fragment_chain:1"
22445..36631
/note="assembly_fragment:04629
fragment_chain:1"
36732..63941
/note="assembly_fragment:02431
fragment_chain:1"
64042..76959
/note="assembly_fragment:06652
fragment_chain:2"
77060..80361
/note="assembly_fragment:02856
fragment_chain:2"
80462..84581
/note="assembly_fragment:05282
fragment_chain:2"
84682..89169
/note="assembly_fragment:01199
fragment_chain:2"

misc_feature
89270..94454
/note="assembly_fragment:00854
fragment_chain:3"
94555..99344
/note="assembly_fragment:04769
fragment_chain:3"
99445..105364
/note="assembly_fragment:03592
fragment_chain:3"
105465..112857
/note="assembly_fragment:02797
fragment_chain:4"
112958..116353
/note="assembly_fragment:05253
fragment_chain:4"
116454..119029
/note="assembly_fragment:01663
fragment_chain:4"
119130..157992
/note="assembly_fragment:03653
fragment_chain:5"
158093..164601
/note="assembly_fragment:05568
fragment_chain:5"
164702..195901
/note="assembly_fragment:01977
fragment_chain:5"
196002..198807
/note="assembly_fragment:00023
fragment_chain:6"
198908..204604
/note="assembly_fragment:05254
fragment_chain:6"
204705..207838
/note="assembly_fragment:00245
fragment_chain:7"
207939..221335
/note="assembly_fragment:03587
fragment_chain:7"
221436..224577
/note="assembly_fragment:00642
fragment_chain:8"
224678..228248
/note="assembly_fragment:04656
fragment_chain:8"
228349..237923
/note="assembly_fragment:01666
fragment_chain:9"
238024..242134
/note="assembly_fragment:05529
fragment_chain:9"
242235..250341
/note="assembly_fragment:02430
fragment_chain:10"
250442..273524
/note="assembly_fragment:00629
fragment_chain:10"
273625..277125
/note="assembly_fragment:04232
fragment_chain:11"
277226..282694
/note="assembly_fragment:02725
fragment_chain:11"
282795..284888
/note="assembly_fragment:04893
fragment_chain:12"
284989..290372
/note="assembly_fragment:01882
fragment_chain:12"
290473..298809
/note="assembly_fragment:00742"
298910..301400
/note="assembly_fragment:00762"

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10161
Center clone name: 660_H_19

* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

| | | |
|---|-------|-----------------------------------|
| 1 | 685: | contig of 685 bp in length |
| * | 686 | 785: gap of 100 bp |
| * | 786 | 1513: contig of 728 bp in length |
| * | 1514 | 1613: gap of 100 bp |
| * | 1614 | 2354: contig of 741 bp in length |
| * | 2355 | 2454: gap of 100 bp |
| * | 2455 | 3184: contig of 730 bp in length |
| * | 3185 | 3284: gap of 100 bp |
| * | 3285 | 3959: contig of 675 bp in length |
| * | 3960 | 4059: gap of 100 bp |
| * | 4060 | 4746: contig of 687 bp in length |
| * | 4747 | 4846: gap of 100 bp |
| * | 4847 | 5564: contig of 718 bp in length |
| * | 5565 | 5664: gap of 100 bp |
| * | 5665 | 6360: contig of 696 bp in length |
| * | 6361 | 6460: gap of 100 bp |
| * | 6461 | 7132: contig of 672 bp in length |
| * | 7133 | 7232: gap of 100 bp |
| * | 7233 | 7937: contig of 705 bp in length |
| * | 7938 | 8037: gap of 100 bp |
| * | 8038 | 8749: contig of 712 bp in length |
| * | 8750 | 8849: gap of 100 bp |
| * | 8850 | 9522: contig of 673 bp in length |
| * | 9523 | 9622: gap of 100 bp |
| * | 9623 | 10356: contig of 734 bp in length |
| * | 10357 | 10456: gap of 100 bp |
| * | 10457 | 11181: contig of 725 bp in length |
| * | 11182 | 11281: gap of 100 bp |
| * | 11282 | 11975: contig of 694 bp in length |
| * | 11976 | 12075: gap of 100 bp |
| * | 12076 | 12785: contig of 710 bp in length |
| * | 12786 | 12885: gap of 100 bp |
| * | 12886 | 13557: contig of 672 bp in length |
| * | 13558 | 13657: gap of 100 bp |
| * | 13658 | 14347: contig of 690 bp in length |
| * | 14348 | 14447: gap of 100 bp |
| * | 14448 | 15132: contig of 685 bp in length |
| * | 15133 | 15232: gap of 100 bp |
| * | 15233 | 15912: contig of 680 bp in length |
| * | 15913 | 16012: gap of 100 bp |
| * | 16013 | 16731: contig of 719 bp in length |
| * | 16732 | 16831: gap of 100 bp |
| * | 16832 | 17506: contig of 675 bp in length |
| * | 17507 | 17606: gap of 100 bp |
| * | 17607 | 18339: contig of 733 bp in length |

* 1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 858 1635: contig of 778 bp in length
* 1636 1735: gap of 100 bp
* 1736 2490: contig of 755 bp in length
* 2491 2590: gap of 100 bp
* 2591 3343: contig of 753 bp in length
* 3344 3443: gap of 100 bp
* 3444 4224: contig of 781 bp in length
* 4225 4324: gap of 100 bp
* 4325 5088: contig of 764 bp in length
* 5089 5188: gap of 100 bp
* 5189 5968: contig of 780 bp in length
* 5969 6068: gap of 100 bp
* 6069 6844: contig of 776 bp in length
* 6845 6944: gap of 100 bp
* 6945 7701: contig of 757 bp in length
* 7702 7801: gap of 100 bp
* 7802 8573: contig of 772 bp in length
* 8574 8673: gap of 100 bp
* 8674 9430: contig of 757 bp in length
* 9431 9530: gap of 100 bp
* 9531 10295: contig of 765 bp in length
* 10296 10395: gap of 100 bp
* 10396 11139: contig of 744 bp in length
* 11140 11239: gap of 100 bp
* 11240 11988: contig of 749 bp in length
* 11989 12088: gap of 100 bp
* 12089 12874: contig of 786 bp in length
* 12875 12974: gap of 100 bp
* 12975 13750: contig of 776 bp in length
* 13751 13850: gap of 100 bp
* 13851 14638: contig of 788 bp in length
* 14639 14738: gap of 100 bp
* 14739 15520: contig of 782 bp in length
* 15521 15620: gap of 100 bp
* 15621 16381: contig of 761 bp in length
* 16382 16481: gap of 100 bp
* 16482 17248: contig of 767 bp in length
* 17249 17348: gap of 100 bp
* 17349 18123: contig of 775 bp in length
* 18124 18223: gap of 100 bp
* 18224 18977: contig of 754 bp in length
* 18978 19077: gap of 100 bp
* 19078 19860: contig of 783 bp in length
* 19861 19960: gap of 100 bp
* 19961 20732: contig of 772 bp in length
* 20733 20832: gap of 100 bp
* 20833 21590: contig of 758 bp in length
* 21591 21690: gap of 100 bp
* 21691 22446: contig of 756 bp in length
* 22447 22546: gap of 100 bp
* 22547 23333: contig of 787 bp in length
* 23334 23433: gap of 100 bp
* 23434 24212: contig of 779 bp in length
* 24213 24312: gap of 100 bp
* 24313 25093: contig of 781 bp in length
* 25094 25193: gap of 100 bp
* 25194 25958: contig of 765 bp in length
* 25959 26058: gap of 100 bp
* 26059 26839: contig of 781 bp in length
* 26840 26939: gap of 100 bp
* 26940 27705: contig of 766 bp in length
* 27706 27805: gap of 100 bp
* 27806 28568: contig of 763 bp in length
* 28569 28668: gap of 100 bp
* 28669 29419: contig of 751 bp in length
* 29420 29519: gap of 100 bp
* 29520 30311: contig of 792 bp in length
* 30312 30411: gap of 100 bp
* 30412 31202: contig of 791 bp in length
* 31203 31302: gap of 100 bp
* 31303 32082: contig of 780 bp in length

* 32083 32182: gap of 100 bp
* 32183 32934: contig of 752 bp in length
* 32935 33034: gap of 100 bp
* 33035 33796: contig of 762 bp in length
* 33797 33896: gap of 100 bp
* 33897 34683: contig of 787 bp in length
* 34684 34783: gap of 100 bp
* 34784 35552: contig of 769 bp in length
* 35553 35652: gap of 100 bp
* 35653 36406: contig of 754 bp in length
* 36407 36506: gap of 100 bp
* 36507 37294: contig of 788 bp in length
* 37295 37394: gap of 100 bp
* 37395 38159: contig of 765 bp in length
* 38160 38259: gap of 100 bp
* 38260 39017: contig of 758 bp in length
* 39018 39117: gap of 100 bp
* 39118 39890: contig of 773 bp in length
* 39891 39990: gap of 100 bp
* 39991 40774: contig of 784 bp in length
* 40775 40874: gap of 100 bp
* 40875 41643: contig of 769 bp in length
* 41644 41743: gap of 100 bp
* 41744 42521: contig of 778 bp in length
* 42522 42621: gap of 100 bp
* 42622 43393: contig of 772 bp in length
* 43394 43493: gap of 100 bp
* 43494 44288: contig of 775 bp in length
* 44289 44388: gap of 100 bp
* 44389 45149: contig of 781 bp in length
* 45150 45249: gap of 100 bp
* 45250 46024: contig of 775 bp in length
* 46025 46124: gap of 100 bp
* 46125 46868: contig of 744 bp in length
* 46869 46968: gap of 100 bp
* 46969 47720: contig of 752 bp in length
* 47721 47820: gap of 100 bp
* 47821 48596: contig of 776 bp in length
* 48597 48696: gap of 100 bp
* 48697 49462: contig of 766 bp in length
* 49463 49562: gap of 100 bp
* 49563 50335: contig of 773 bp in length
* 50336 50435: gap of 100 bp
* 50436 51207: contig of 772 bp in length
* 51208 51307: gap of 100 bp
* 51308 52078: contig of 771 bp in length
* 52079 52178: gap of 100 bp
* 52179 52954: contig of 776 bp in length
* 52955 53054: gap of 100 bp
* 53055 53813: contig of 759 bp in length
* 53814 53913: gap of 100 bp
* 53914 54675: contig of 762 bp in length
* 54676 54775: gap of 100 bp
* 54776 55537: contig of 762 bp in length
* 55538 55637: gap of 100 bp
* 55638 56413: contig of 776 bp in length
* 56414 56513: gap of 100 bp
* 56514 57278: contig of 765 bp in length
* 57279 57378: gap of 100 bp
* 57379 58149: contig of 771 bp in length
* 58150 58249: gap of 100 bp
* 58250 59023: contig of 774 bp in length
* 59024 59123: gap of 100 bp
* 59124 59906: contig of 783 bp in length
* 59907 60006: gap of 100 bp
* 60007 60792: contig of 786 bp in length
* 60793 60892: gap of 100 bp
* 60893 61663: contig of 771 bp in length
* 61664 61763: gap of 100 bp
* 61764 62539: contig of 776 bp in length
* 62540 62639: gap of 100 bp
* 62640 63422: contig of 783 bp in length

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-660H19
Unpublished
2 (bases 1 to 54398)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galaquán, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mhova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10161
Center clone name: 660_H_19

* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 685: contig of 685 bp in length
* 586 785: gap of 100 bp
* 786 1513: contig of 728 bp in length
* 1514 1613: gap of 100 bp
* 1614 2354: contig of 741 bp in length
* 2355 2454: gap of 100 bp
* 2455 3184: contig of 730 bp in length
* 3185 3284: gap of 100 bp
* 3285 3959: contig of 675 bp in length
* 3960 4059: gap of 100 bp
* 4060 4746: contig of 687 bp in length
* 4747 4846: gap of 100 bp
* 4847 5564: contig of 718 bp in length
* 5565 5664: gap of 100 bp
* 5665 6360: contig of 696 bp in length
* 6361 6460: gap of 100 bp
* 6461 7132: contig of 672 bp in length
* 7133 7232: gap of 100 bp
* 7233 7937: contig of 705 bp in length
* 7938 8037: gap of 100 bp
* 8038 8749: contig of 712 bp in length
* 8750 8849: gap of 100 bp
* 8850 9522: contig of 673 bp in length
* 9523 9622: gap of 100 bp
* 9623 10356: contig of 734 bp in length
* 10357 10456: gap of 100 bp
* 10457 11181: contig of 725 bp in length
* 11182 11281: gap of 100 bp
* 11282 11975: contig of 694 bp in length
* 11976 12075: gap of 100 bp
* 12076 12785: contig of 710 bp in length
* 12786 12885: gap of 100 bp
* 12886 13557: contig of 672 bp in length
* 13558 13657: gap of 100 bp
* 13658 14347: contig of 690 bp in length
* 14348 14447: gap of 100 bp
* 14448 15132: contig of 685 bp in length
* 15133 15232: gap of 100 bp
* 15233 15912: contig of 680 bp in length
* 15913 16012: gap of 100 bp
* 16013 16731: contig of 719 bp in length
* 16732 16831: gap of 100 bp
* 16832 17506: contig of 675 bp in length
* 17507 17606: gap of 100 bp
* 17607 18339: contig of 733 bp in length
* 18340 18439: gap of 100 bp
* 18440 19109: contig of 670 bp in length
* 19110 19209: gap of 100 bp
* 19210 19900: contig of 691 bp in length
* 19901 20000: gap of 100 bp
* 20001 20693: contig of 693 bp in length
* 20694 20793: gap of 100 bp
* 20794 21464: contig of 671 bp in length
* 21465 21564: gap of 100 bp
* 21565 22245: contig of 681 bp in length
* 22246 22345: gap of 100 bp
* 22346 23049: contig of 704 bp in length
* 23050 23149: gap of 100 bp
* 23150 23857: contig of 708 bp in length
* 23858 23957: gap of 100 bp
* 23958 24685: contig of 728 bp in length
* 24686 24785: gap of 100 bp
* 24786 25499: contig of 714 bp in length
* 25500 25599: gap of 100 bp
* 25600 26302: contig of 703 bp in length
* 26303 26402: gap of 100 bp
* 26403 27108: contig of 706 bp in length
* 27109 27208: gap of 100 bp
* 27209 27910: contig of 702 bp in length
* 27911 28010: gap of 100 bp
* 28011 28720: contig of 710 bp in length
* 28721 28820: gap of 100 bp
* 28821 29484: contig of 664 bp in length
* 29485 29584: gap of 100 bp
* 29585 30328: contig of 744 bp in length
* 30329 30428: gap of 100 bp
* 30429 31111: contig of 683 bp in length
* 31112 31211: gap of 100 bp
* 31212 31931: contig of 720 bp in length
* 31932 32031: gap of 100 bp
* 32032 32710: contig of 679 bp in length
* 32711 32810: gap of 100 bp
* 32811 33502: contig of 692 bp in length
* 33503 33602: gap of 100 bp
* 33603 34301: contig of 699 bp in length
* 34302 34401: gap of 100 bp
* 34402 35100: contig of 699 bp in length
* 35101 35200: gap of 100 bp
* 35201 35914: contig of 714 bp in length
* 35915 36014: gap of 100 bp
* 36015 36708: contig of 695 bp in length
* 36710 36809: gap of 100 bp
* 36810 37518: contig of 709 bp in length
* 37519 37618: gap of 100 bp
* 37619 38330: contig of 712 bp in length


```

QY 5026 ttgtgaataagtcgaagttcttggaattactttacacctcagtaggtattgtccccaga 5085
|||||
Db 26154 TTGTGAATAAGTCCTGAAGTCTCTGGATTACTTTACACCTCAGTATTGATTGTGCCAGA 26213
|||||
QY 5086 attttctggccttctcatggaataaataaattttaagaagaagaattttaagaattttaatt 5145
|||||
Db 26214 ATTTCTGGCCTTCTCATGCGCAATGCAAAATTTTAAGAGAAGAAGATTTAAGATATTTAAAT 26273
|||||
QY 5146 ttaaaagtggtgtataaaataatgactgaattcttcttccatttccatctcttca 5205
|||||
Db 26274 TTAAGAGTGTGTATATAATAATGACTGCAATCTTTATCCCATTTTATCATCTTTCA 26333
|||||
QY 5206 gtttttattactactgtatcaataaaaaattctgttaatttgaatgagtaaaaaa 5265
|||||
Db 26334 GTTTTATTAACTACTGTATCAATAAAATTCGTAAATTTGAATGAGTTTAAATAGTCT 26393
|||||
QY 5266 aaaa 5269
|||||
Db 26394 AGAA 26397
|||||

RESULT 8
AC026889 2079 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ23236 fis, clone COL00725.
AC026889 AK026889
VERSION AK026889.1 GI:10439854
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL00725.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2079)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@lms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
Source Location/Qualifiers
1. .2079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL00725"
/clone_lib="COL"
/tissue_type="colon"
/notes="cloning vector pME18SFL3"
304. .681
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15584.1"
/db_xref="GI:10439855"
/translation="MDDLTKLVQEQPKGSRKRGHTASEDSDEQMPPEERKLKEDI
LNEDEONPPPKKRGKRPKPPLGGTTPKEPTMTSKKSKKKSGPPAPEEEEEER
QSGNTEQSKSKQHRVSRRAQR"
BASE COUNT 759 a 354 c 339 g 627 t

```

ORIGIN

```

Query Match 10.7%; Score 566.4; DB 89; Length 2079;
Best Local Similarity 87.7%; Pred. No. 7.1e-96;
Matches 670; Conservative 0; Mismatches 1; Indels 93; Gaps 1;

QY 3689 gtctgaattggagaagcctagagcaggaagaaaccccgctcacagaacagaggagaa 3748
|||||
Db 237 GTCTGAATTGCGAAGCCTAGAGCGCAGGAGAAACCCCGTCACAGAACAGGAGGAGAA 296
|||||
QY 3749 attaggtatgtagtgaacttgactaagttggtacaggaacagaacacctaagagcagcag 3808
|||||
Db 297 ATTAGGTATGATGACTTGACTTAAGTTGGTACAGGACAGAAACCTTAAGGCGATCAGCG 356
|||||
QY 3809 aagtcggaagaaagggccatcacgcttcagaaatctgtgaacacagcagtggtgaggaaaa 3868
|||||
Db 357 AGTCGGAAAGAGGCCATACGGCTTCAGAACTCTGTATGACACAGCATGGCTTGAGGAAA 416
|||||
QY 3869 gaggtctcaagaagatatattagaataatgaagatgaacagaatagtcgcgcaaaaaggg 3928
|||||
Db 417 GAGGCTCAAGAAGATATATTAGAAAAATGAAGATGAACAGATAGTCGCCCAAAAAAGGG 476
|||||
QY 3929 taaaagagggcgcaccacccaaaacacctctgtgaggtgtacaccacaaagaaagagcaacaat 3988
|||||
Db 477 TAAAGAGGCGCGCACCAACCAAAACCTCTTGTGGAGGTACACCAAAAGAGAGCCACAAT 536
|||||
QY 3989 gaaaactcttaaaaaaggaagcaaaaaaaatctggacctccagcaccagagagaggagga 4048
|||||
Db 537 GAAACTCTTAAAAAGGAAGCAAAAAAAATCTTGGACCTCCAGCACCACAGAGGAGGAGGA 596
|||||
QY 4049 agaaagaaagacaaagtggaaatcacggacagagagtcgaaagcaaaagcaaacagcagcag 4108
|||||
Db 597 AGAAGAAGAAAGACAAAGTGGAAATACGGGAACAGAAAGTCCAAAGCAACAGCACCAGT 656
|||||
QY 4109 gtaagagagagacacagcag----- 4127
|||||
Db 657 GTCAAGGAGAGCAGCAGCAGAGGTAAAGCATGTGTAACCTCTAAACTGTCATCTGTTTCGTAC 716
|||||
QY 4128 -----agagcaga 4135
|||||
Db 717 TATATTATAAATCATAATTGTGATGCTATCCACATTTGGGTCTTCCCAAGAGCAGAGCAGA 776
|||||
QY 4136 atctctgaatctagtgcaattgaatccacacagtcacacacacagagaagagagagag 4195
|||||
Db 777 ATCTCTGAATCTAGTGCATTTGAATCCACACAGTCACACACACAGAGGAGGAGGAG 836
|||||
QY 4196 accatcaaaaccccatcaccatcaccaacaaacaaataatgttaagttaataattac 4255
|||||
Db 837 ACCATCAAAACCCCATCACCATCACACCAACCAAAAAAATGTGTAGTTGTAATATTAC 896
|||||
QY 4256 attcaaaccaatttcaaatattttgcaaaagtctcctaatttgaacatacatattg 4315
|||||
Db 897 ATTTCAACCAATTTCAAAATATTTCGCAAAAGTTCCTTAATTTGTAACATACATATTG 956
|||||
QY 4316 ctgtattaaattccatatatttagcccccattacactaggtacg 4359
|||||
Db 957 CTGTATTAAATTCATATATTATTAGCCCCATTACACTAGGTAAG 1000
|||||

RESULT 9
AC068224 54398 bp DNA HTG 30-APR-2000
LOCUS Homo sapiens chromosome 3 clone RP11-660H19 map 3, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC068224 AC068224
VERSION AC068224.1 GI:7671284
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 54398)

```



```

72405; Paired with EST T51543 matching this clone; match
H91918"
repeat_region 28327..28360
/notes="17 copies of 2 mer 88 % conserved"
prim_transcript <28864..>25013
/notes="match: 3' EST T51543 clone 72405; Paired with EST
T51695 matching this clone"
29300..29456
repeat_region /note="MIR2 repeat: matches 146..1 of consensus"
30648..30934
repeat_region /note="AluSc repeat: matches 7..302 of consensus"
33156..33439
repeat_region /note="AluSc repeat: matches 301..2 of consensus"
34900..35082
repeat_region /note="L1MD2 repeat: matches 425..263 of consensus"
34982..35061
repeat_region /note="L1ME3 repeat: matches 343..264 of consensus"
37923..38245
repeat_region /note="AluSc repeat: matches 294..1 of consensus"
38457..38536
repeat_region /note="MIR repeat: matches 61..161 of consensus"
39265..39568
repeat_region /note="AluJo repeat: matches 302..3 of consensus"
40952..41046
repeat_region /note="MIR repeat: matches 29..125 of consensus"
43042..43331
repeat_region /note="AluJo repeat: matches 301..11 of consensus"
43583..43669
repeat_region /note="MER5A repeat: matches 10..99 of consensus"
43881..44083
repeat_region /note="MIR repeat: matches 248..28 of consensus"
44772..45170
repeat_region /note="MIR repeat: matches 388..1 of consensus"
45848..45949
repeat_region /note="MIR repeat: matches 144..41 of consensus"
46073..46373
repeat_region /note="AluSp repeat: matches 298..1 of consensus"
49357..49436
repeat_region /note="MIR repeat: matches 198..114 of consensus"
50023..50324
repeat_region /note="AluSc repeat: matches 1..302 of consensus"
50826..50892
repeat_region /note="L1MB8 repeat: matches 853..921 of consensus"
51407..51534
repeat_region /note="L1 repeat: matches 4869..4999 of consensus"
51563..51863
repeat_region /note="AluY repeat: matches 1..301 of consensus"
51869..52077
repeat_region /note="AluB repeat: matches 86..298 of consensus;
incomplete repeat"
52311..52712
repeat_region /note="L1ME3A repeat: matches 44..475 of consensus"
52698..52781
repeat_region /note="L1MD2 repeat: matches 569..653 of consensus"
52719..52833
repeat_region /note="L1ME3 repeat: matches 591..707 of consensus"
53752..54490
repeat_region /note="L1 repeat: matches 4668..5390 of consensus"
54313..54870
repeat_region /note="L1MB4 repeat: matches 1..540 of consensus"
55008..55520
repeat_region /note="L1MA9 repeat: matches 530..1055 of consensus"
57267..57487
repeat_region /note="AluY repeat: matches 79..298 of consensus;
incomplete repeat"
57510..57576
repeat_region /note="MIR repeat: matches 76..143 of consensus"
58670..58699
repeat_region /note="15 copies of 2 mer 87 % conserved"
58676..58699
repeat_region /note="6 copies of 4 mer 100 % conserved"
59808..60115
repeat_region /note="L1MB7 repeat: matches 895..571 of consensus"

```

```

60135..60526
repeat_region /note="MIR repeat: matches 505..108 of consensus"
60531..60815
repeat_region /note="AluB repeat: matches 2..290 of consensus"
60816..60915
repeat_region /note="MIR repeat: matches 114..3 of consensus"
60918..61113
repeat_region /note="L1MB6 repeat: matches 529..320 of consensus"
61115..61379
repeat_region /note="AluY repeat: matches 299..31 of consensus;
incomplete repeat"
61380..61675
repeat_region /note="L1MB5 repeat: matches 309..1 of consensus"
61540..61805
repeat_region /note="L1 repeat: matches 5389..5107 of consensus"
61916..62212
repeat_region /note="AluSc repeat: matches 1..299 of consensus"
64361..64420
repeat_region /note="U6 repeat: matches 63..1 of consensus"
64816..65101
repeat_region /note="AluJo repeat: matches 2..302 of consensus"
65202..65856
repeat_region /note="L1PA16 repeat: matches 902..217 of consensus"
66393..66470
repeat_region /note="L1 repeat: matches 2101..2178 of consensus"
66530..66868
repeat_region /note="L1 repeat: matches 2265..2610 of consensus"
66881..67104
repeat_region /note="MER30 repeat: matches 2..230 of consensus"
67104..67855
repeat_region /note="L1 repeat: matches 2613..3397 of consensus"
67999..68677
repeat_region /note="L1 repeat: matches 4255..4942 of consensus"
68693..68972
repeat_region /note="AluSc repeat: matches 299..1 of consensus"
68983..69436
repeat_region /note="L1 repeat: matches 4954..5390 of consensus"
69283..70242
repeat_region

Query Match 17.1%; Score 901.6; DB 92; Length 137246;
Best Local Similarity 96.0%; Pred. No. 1.9e-158;
Matches 925; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 4306 atacatatgtctgtatttaataatccatatttagccccattacactagtagtcagcg9g9 4365
Db 25955 ATGCACATATACCCATTTTAATTAAGCATCTGTGCTTTCCCTTTTAAGGTACGGCGGCG 26014
QY 4366 aagtgtctaaagggaacgcatgaacaaatgaataataacttctctgtgaaagctt 4425
Db 26015 AAGTGTCTAAAGGAACGGCGATGAACAAATGAATTAATAACTTTCTCTGTGAAAGCTT 26074
QY 4426 tggaaaaatcttttttttttttttttttttttttttttttttttttttttttttttttttt 4485
Db 26075 TGGAAAAATCTTTTtttttttttttttttttttttttttttttttttttttttttttttt 26134
QY 4486 ttagcacaaaaatgggactgctgaagtgacagcttgacacttacttttggtagcccat 4545
Db 26135 TGATGCACAAAAATGGGACTGCTGAAGATGGACAGTTGGACCTTACTTTTGGTGACCCCAT 26194
QY 4546 acatttgtgtcacatgcttttagccatcacatggaacattgaactatgagcttctgtga 4605
Db 26195 ACATTTGTGTGCATGCTTTAGCCATACACATGGTAACATTGACATGAGTCTTGTGA 26254
QY 4606 aagtgttaatgtcgatggctgtatgtacataaagaagaacttggtaaatcttttttct 4665
Db 26255 AAGTGAATGTGGATGGCTATGTACACATAAAGAAACATTTGTAATATCTTTTCT 26314
QY 4666 ttttttaattgtctgtattctgaagtgtctgtatagtttttctgtcggtttaaact 4725
Db 26315 TTTTtttttaattgtctgtattctgaagtgtctgtatagtttttctgtcggtttaaact 26374
QY 4726 gacagtaccgactgtttattgattctattgattgaagaatttggtaggatact 4785

```



```

REFERENCE
AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@fokazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
REFERENCE
AUTHORS Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro
JOURNAL DNA Res. 5 (3), 169-176 (1998)
MEDLINE 98403880
FEATURES
Source
1..5177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HJ03994"
/cclone_lib="pBluescriptII SK plus"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
1..2558
/gene="KIAA0648"
<1..2558
/gene="KIAA0648"
/codon_start=3
/product="KIAA0648 protein"
/protein_id="BAA31623.1"
/db_xref="GI:3327110"
/translation="LYLYASLDPNNAKALNEMKCNMLRSHVRELLDLHKQPTSEA
NCSEMGKLMITAKNLPDPKGAODPVKFNQVLGDDKLRSLQLELLISPTCSCKOADI
CYRETFARKLPKQPTNPLEMKVLELLERLAPVHIDSEALSLVLMKNSIEGTADDE
EGVSPDRAISGLLELLKVSFTHTPSFSAETYESLQCLRWEDDVAERAAQIFRN
TGKRLTETDLPQIRSLTIPILHOKAKRQTPHOAKOVAHCIIAIFTNKEVQLAQIFPELS
RSLNADVPQELTPIVLGHSMLAPQDPQAFSPKSVVAFIVKDLLMNDRTGKNGK
LWSPDEVSPEVLAKQAIKILIRWLLGMKNQSKNSANSTLRLLSAMLVSEGDITPEOK
RISKDSMRSLRQSAITMKLAOEPYHEIITPEQFOLCALVINDECVQVRIFAOKL
HKALVKLLPVMYAFALCAKDPVKERRAHARQCLLNISIRRYIKONPMATEKLL
SLLPEYVPMYHLLAHADPDFTRSQVDLQRIKELMFLMVEVLMKNNESHAPMKK
MAENIKLRDAQSPDESTNEKLYTCDVALCVINSKSALCNADSPKQVLPMPKEFTQ
PEKDCNDKSYISEETRVLLTGKPKPAGVLGNKPLSATGRKPYVRSITGETGNSI
NYNSLNFSTGSRSEQSEAEATGVSENEENPVRIISVTPVKNTIDPVKNKEINSDQA
TCGNISSTGRGKRRTVTAAGAENIQOKTDEKVDSEGGPPAPSKPRGRRPKSKSQGNATK
NDDLANKPTNKGKRAAVGOESPGGLEAGNAPKPLQDLAKKAAPAEROIDLQR"
BASE COUNT 1690 a 931 c 1037 g 1519 t
ORIGIN

Query Match 19.8%; Score 1041.4; DB 85; Length 5177;
Best Local Similarity 68.0%; Pred. No. 1.6e-184;
Matches 1450; Conservative 0; Mismatches 681; Indels 0; Gaps 0;

QY 1492 gcttattacttgatgacacactgatttaaatgctgtgaagcattgaatgaaatg 1551
Dbb 1 GCTTATTACTTATATAGTCTTGGATCCAAATGCTGTAAAGCTCTCAACGAAATGT 60

QY 1552 ggaattgtcaaatctgctccgacatcaagataaagatttgcttgactgattaagcaac 1611
Dbb 61 GGAAGTGTGAGAACATGCTTCGGAGCGATGATCGCGCAAGATATTGGATTTGCAACAGCAGC 120

QY 1612 ccaaacacagatgcagatgtcaagggccattatttcaaaagtgtggttattacaagaatt 1671
Dbb 121 CTACATCAGAGGCTACTGTTCTGCCATGTTTGGAAACATGATGACCATAGCAACAAT 180

QY 1672 taactgatccttgtaagggctcaggattcatatgaagaaatttcacacaggtgttagaagatg 1731
Dbb 181 TGCGTCACCCCGGAAAGCACAGATTTTGTGAAGAAATTTAACACAGGTTCTCGGGGATG 240

QY 1732 atgagaaaataagaagacaggttagaagtagtctgttagtccaacatgctcctgcagcagg 1791

```

```

Db 241 ATCAGAAACTTCGGTCTCAGTTGGAGTTATTAAATTAGCCCAACCTGTTCTTGCAAAAG 300
QY 1792 ctgaagagtgctggtgaaataaagaagttgggcaaccccaacacagcctcacaatc 1851
Db 301 CAGATATTGTGTGAGAGAAATAGCCCGGAAACTTCGAAATCTTAAGCAACCAACAAAT 360
QY 1852 ctttcctggaatgatcaagttctctcttgagagagtagcacctgtgcacatagatagcag 1911
Db 361 CTTTTCTCAGAGTGGTCAAAATTTCTGTTGGAAGAAATCGCACCTGTGCACATTTGATTG 420
QY 1912 aatctatcagtgctcttataaacaagtgaaacaaatcaatagatggaacagcagatgag 1971
Db 421 AAGCCATAAGTGCACCTAGTGAATTTGATGAATAAGTCAATAGAGGGGACAGCAGATGATG 480
QY 1972 aagatgaggggtgtccaaactgatccaaagccatcagagcaggtcttgactgcttaaggtac 2031
Db 481 AAGAGGAGGGTGTAGTCCAGATACAGCTATCGGTTTCAGAGCTTGAACTCTTTAAGGTT 540
QY 2032 tctctttacacatcccatctcattcttcttgctggaacatttgatcattactgcttgcct 2091
Db 541 TGTCTTTTACACATCTACCTCGTTCACCTCTGCAGAGACATATGAGTCTCTTGTTCAGT 600
QY 2092 gttcgaagatgatgaaaaagtagcagaagctgacacacacacacacacacacacag 2151
Db 601 GCCTAAGAATGGAGGATGACAGGTAGCAGAGCTGCTATTCAAAATTTTGTAGAAATACAG 660
QY 2152 gaagcaaatgtgaagaggtatttccacacatcagatcagcagcttgcctctgcttgcctc 2211
Db 661 GTCACAAATAGAACAGACCTTCCCCAGATAGCATCGACCTTAATCCCATTTTACATC 720
QY 2212 acaaatctaaaaagagcccccccgctcaagccaaatgatgccattctgtatccatgcga 2271
Db 721 AAAAGCAAGAGGGGTACTCCACACCAAGAAACAGGCTGTGCACTGTATACACGCCA 780
QY 2272 tatttctagtaaaagacccagtttgacacagatatttgagcctctgcataagagcctag 2331
Db 781 TATTCACAAATAAAGAAAGTCCAGCTTTCACAGATTTTTCAGCCACTTCAGTAGAGTCTGA 840
QY 2332 atcccaagcaacctggaacatctcataaacaccattggttactattgtgctcataatgctctcc 2391
Db 841 ATGCTGATGTGCCAGAACAACTTATAACTCCATAGTTTCATTTGGGCCACATTTCTATGT 900
QY 2392 ttgcaactgatacaatttgctgctctcttggaatcttggttagctactcttcattgtgaaag 2451
Db 901 TAGCACAGATCAGTTTGTCTTCCCAATGAAATCTCTAGTAGCAAAATTTTATTGTGAAAG 960
QY 2452 atcttctgatgaatcggttccaggggaaagaaagacaaactaaactttgggttccagatg 2511
Db 961 ATCTGCTAATGAATGACAGGTCAACAGGTGAAAGAAAGTGGAAACATGTGCTTCCAGATG 1020
QY 2512 aagaagtatctcctgagacaaatggtcaaaattcaggctattaaaaatgaggttcgatggc 2571
Db 1021 AAGAGGTTTCCCTTGAAGTACTAGCAAGAGGTACAGGCAATTAACACTTCTGGTAAGGTGC 1080
QY 2572 tacttggaaatgaaaaataatcacagtaaatcagggaactctcactcacttaagattgctaaca 2631
Db 1081 TGTGGGTATGAAAAACAAACACAGTCTAAATCTGCCAATTCGCAACCTTCGCTGTTATTATCAG 1140
QY 2632 caaatctgatgtagtgagacttgacagacaggggaaaaatagtaaacacagatatgt 2691
Db 1141 CGATGTTGTTAGTAGGGGTGACCTGACAGAGCAAGAAAGAGGATCAGTAATCTGATATGT 1200
QY 2692 cagctctgagacttgctgctgggagtgctatgtgagctggtgcacagaaacccctgttac 2751
Db 1201 CTCGCTTGCAGATTAGCTGCTGTTAGTGCATTAATGAAGCTTCTCAGGAACCTTGTACC 1260
QY 2752 atgaataatcacattagaaacaaatcacagctatgtgcatagctatcaacagatgaatgct 2811
Db 1261 ATGAAATTTATACCCCGAGAACAGTTTTCAGCTCTGTGACATTTGTTATTAAATGAGTGT 1320
QY 2812 atcaagtaagacaggttttggccgagaaacttcacaaagcctttcccgcttccagcttc 2871

```


BASE COUNT 609 a 348 c 379 g 505 t 11 others
ORIGIN

Query Match 33.3%; Score 1754.4; DB 97; Length 1852;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1817; Conservative 10; Mismatches 21; Indels 6; Gaps 5;

QY 1129 ttaaaagtgaagtcacatgacccctgaggaagcattatgacatgatgttatgtgtcaatag 1188
Db 1 TTAAGTGAGGTGCATGACCCCTGAGGAAGCTATTGACATGATGTTATTGTGTCAATAG 60

QY 1189 ttacagctgctaaaaagatattctctggtcgaatgacacttacttaatttttgtgagag 1248
Db 61 TTACAGCTGCTAAAGAGATATCTCTGTCATGATGATCTACTTAAATTTGTGAGAG 120

QY 1249 agagaacattagacaaacgatgagagctacgaaagacatgagcctgagcctggccaaa 1308
Db 121 AGAGAACATTAGACAAACGATGGAGAGTACGCAAAAGACCATGATGGGACTTGCCCAAA 180

QY 1309 ttataagaataatgctttcacagtcacagcagctggaagaagcctgcaaaacagatagcat 1368
Db 181 TTTATAAGAAATATGCTTTACAGTCACGACGCTGGAAGAGATGCTGCAAAACAGATAGCAT 240

QY 1369 ggatcaagaacaaattgctacatatattatcaaaatagattgtagtgatgcactcttg 1428
Db 241 GGATCAAGACAAATGCTTACATATATATATCAAAATAGTATTGATGATCGACTACTTG 300

QY 1429 ttgaacggattcttgcctcaatcatggtctcctcacaaatttagaaactacagaacggatga 1488
Db 301 TTGAACGGATCTTTGCTCAATACATGTTCTCTCACATTTTAGAACTTACAGACGGATGA 360

QY 1489 aatgcttatattactgtatgcacacactggatttaaatgctgtgaaagcaattgaatgaaa 1548
Db 361 AATGCTTATATTACTGTATGCCACACTGGATTAAATGCTGTGAAAGCAATTGAATGAAA 420

QY 1549 tgtggaatgtcaaaaactctgctccgacatacaagttaaaggatttgcctgacttgattaaac 1608
Db 421 TGTGGAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGC 480

QY 1609 aacccaaacagatgcagtgctcaaggccatttttcaaaagtgtggttattacaagaa 1668
Db 481 AACCCAAACAGATGCCAGTGTCAAGGCTCATATTTTCAAAAGTGATGTTATTACAAGAA 540

QY 1669 atttacctgactcctgtgaaagctcagatttcattgaagaataattcacacaggtgttagaag 1728
Db 541 ATTTACCTGATPCTTGGTAAAGCTCAGGATTTCAATGAAGAAATTACACACAGGTGTGAAG 600

QY 1729 atgatgagaaaaataagaagcagtagaagtagctttagtcccaacatgctcctgtcaagc 1788
Db 601 ATGATGAGAAAAATAAGAAAGCAGTTAGAGTACTTGTGTTAGTCCAAACATGCTCTGCAAGC 660

QY 1789 aggtcgaaggtgtgtgctgtaaaataactaagaagttgggcaaccccaacacagcctacaa 1848
Db 661 AGGCTGAAGGTGTGTGCGTGAATAACTAAGAAGTTGGGCAACCCCAACAGCCTACAA 720

QY 1849 atccttccctggaatatcatcaagtttctctgtgagagatagacacctgtgcacatagata 1908
Db 721 ATCCTTTTCCGGAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATA 780

QY 1909 ccgaatctacagtgctcttattaaacaagtgaaacaataatcaatagatggaaacagcagatg 1968
Db 781 CCGAATCTATCAGTGCCTTATTAAACAAGTGAACAATCAATAGATGGAAACAGCATG 840

QY 1969 atgaagatgaggggtgttcccaactgatcaagccatcagagcaggtcttgaactgttaagg 2028
Db 841 ATGAAGATGAGGGTGTGTCCAACTCATCAAGCCATCAGACAGGTCTTGAACGTGCTTAAG 900

QY 2029 tactctcatttaacatcccatctcatttcaattctgtgtgaaacaatttgaatcattactgg 2088
Db 901 TACTCTATTACACATCCCATCTCATTTCTTCTGCTGAAACATTTTGAATCATTACTGG 960

QY 2089 cttgtctgaataatggatgataaaaaagtagcagaagctgcactacataaattttcaaaaaca 2148
Db 961 CTTGCTCTGAAATGGATGATGAAAAAGTAGCAGAAGCTGCACATACAAATTTTCAAAAACA 1020

QY 2149 caggaagcaaaatgaagagattttccacacatcagatcagccttctctctctgttttac 2208
Db 1021 CAGGAAGCAAAATGAAGAGATTTTCCACACATCAGATCAGCCTTCTCTCTCTTTTAC 1080

QY 2209 atcacaaatctaaaaagagaccccccg-tcaagcgaataatccattcttatttccat 2267
Db 1081 ATCACAATCTAAAAAGGACCCCCCGCTCAAGCCAAATATGCCATTCATTTGATCCAT 1140

QY 2268 gcgatattttctagtaagagacccagtttgcacagatatattgagcctctgcaaaagc 2327
Db 1141 GCGATATTTTCTAGTAAGAGACYCAGTTTGCACAGATATTGAGCCTCTGCAATAAGC 1200

QY 2328 ctagatccaaagcaacctggaaacatctcataacacctatgggttactatttgggtcatttgc 2387
Db 1201 CTAGATCCAAGCAACCTTGAACATCTCATAACACCATTTGGTTACTATTGGTTCATATTGCT 1260

QY 2388 ctcttgcacctgatcaatttgcctcttggaatcttgggttagctacttcttcttcttctg 2447
Db 1261 CTCTTGCACCTGATCAATTTGCTGCTCTTTGAAATCTTTGGTAGCTACTTTTCATTTG 1320

QY 2448 aaagatcttctcatgaatgatcggtctccaggggaagaaagacaactaaactttgggttcca 2507
Db 1321 AAGATCTTCTCATGATGATCGCTTCCAGGGGAAAGAACAACTAACTTTGGGTTCCA 1380

QY 2508 gatgaagaagtatctctctgagacaatgggtcaaaattcaggctatttaaaatgatggttcga 2567
Db 1381 GATGAAGAAGTATCTCTCGAGACAATGGTCAAAATTCAGGCTATTAAATGATGTTGTCGA 1440

QY 2568 tggctacttgaatgaaaaataataacagtaaaatcaagaacttctacttaag-attgct 2626
Db 1441 TGGCTACTTGGCATGTAAATTAATTCACAGTATATCAGGATCTTCTTCTTAAGSATTTGCT 1500

QY 2627 aacaacaatttgcatagtgatggagacttgacagaacaggggaaatagtaaacaga 2686
Db 1501 AACAAACAATATCGATCGTGTGCGGA-TTGGMAGAMCAGGGGAAA--TTAGTAACCCAGA 1557

QY 2687 tatgtcactgtgagacttgcctgctggagtgcttatttgaagctggtgcacaaagacctg 2746
Db 1558 TATGTCAGCTCTGAGACTTGTCTGCGGAGTGTCTATTGTAAGCTGGCACAAAGAACCTG 1617

QY 2747 ttaccatgaataatcatcattagaaacaataatcagctatgtcattagctatcaacgatga 2806
Db 1618 TTACCATGAATCTCATAWTAGAACCAATATCAGCTATGTCRTTAGCTATCAACAGATGA 1677

QY 2807 atgctatcaagtaagacaagttttt-gcccagaacacttcaaaagcctttcccctttac 2865
Db 1678 ATGCTATCAAGTAAGACAAGTGTGGCCCAAGAACTTCGCAAAAGCCTTTCCCGTTGCG 1737

QY 2866 ggttccacttgagtatatggcaatctgtgccctttdtgcaaaagatcctgtgaagagaga 2925
Db 1738 GGCTTCCACTTGAGAATATGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTGAAGGAGA 1797

QY 2926 gaagagctatgctcagcaatgtttgtgaaaaataataatgtaagcggaggt 2979
Db 1798 GAAGAGCTCATGCTAGGCAATGTTTGGTGAATAATAATGTAAGCGGGACT 1851

RESULT 5

AB014548
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AB014548 5177 bp mRNA
Homo sapiens mRNA for KIAA0648 protein, partial cds.
AB014548.1 GI:3327109
Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:HJ03994.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

| | | | |
|----|------|---|------|
| Qy | 3697 | tggaagaacccatagagcagcagaaataaacccgcgtccacagaaacagagagagaaataggta | 3755 |
| Db | 3541 | TGGAGAACCCCTAGAGCGACGAGAAATAAACGCCCGTCCACAGAACAGGAGGAGAAATTAGGTA | 3600 |
| Qy | 3757 | tgatgacttgactaagtgggtacaggaaacagaacactaaaggcagtcagcgaagtgcga | 3816 |
| Db | 3601 | TGGATGACTTGTACTTAAGTTGGTACAGGAACAGAAACCTTAAGGGCAGTCAGCGAAGTCGGA | 3660 |
| Qy | 3817 | aaagagccatacagcgtctcagaaatctgtgaacagcagtcggcctgagaaagaggtctca | 3876 |
| Db | 3661 | AAAGAGSCCATACGGCTTCAGAAATCTGATGAACACGACGTCGCTGAGGAAAAGAGGCTCA | 3720 |
| Qy | 3877 | aaagaatatattagaaaaatgaagtgaacagaaatagtcgcccaaaaaagggttaaaagag | 3936 |
| Db | 3721 | AAGAAGATATATTAGAAAAATGAAGATGAACAGAAATAGTCCGCCAAAAAAGGGTAAAAAGAG | 3780 |
| Qy | 3937 | qccgaccaccaaacctcttgatggaggtcacaccaaaagaagaccacaatgaaaaactt | 3996 |
| Db | 3781 | GCCGACCACCAAAACCTCTTGTGTGGAGGTACACCAAAAGAAGCGCAACAATGAAAAACTT | 3840 |
| Qy | 3997 | ctaaaaaggaagcacaataaaatctggacctccagccaccagagagaggaagaaagaag | 4056 |
| Db | 3841 | CTRAAAAAAGCAACAAAAAATCTGGACCTCCAGCACAGAGGAGGAGGAGCAAGAAG | 3900 |
| Qy | 4057 | aaagacaaagtcgaataacggaacagaagtcacaaagcaaacagcaccagagtgcaagga | 4116 |
| Db | 3901 | AAAGACAAAGTGGAAATAGCGAACAGAAAGTCCAAAGACAAACAGCACCGAGTGTCAAGGA | 3960 |
| Qy | 4117 | gagcacagagagacagaatctcctgaatctagtgcgaatigaatccacagctccacac | 4176 |
| Db | 3961 | GAGCACAGCAGAGAGCAGAAATCTCTTGAATCTAGTGTCAATTTGAATPCCACACAGTCCACAC | 4020 |
| Qy | 4177 | cacagaagagcagaggaagaccatcaaaaagcccatcaccatcacacacaaaaaaaatg | 4236 |
| Db | 4021 | CACAGAAAGGACGAGGAGAACCATCAAAAAGCCCATCACCATCACACCAAAAAAANAATG | 4080 |
| Qy | 4237 | tgtaagtgtaaataattacatttcaa---accaatttcaaattattttgcaaaagttccct | 4293 |
| Db | 4081 | TCGCTGTAGGACGCTCCAAACAGCAGCTACTTAAGGAAAAATGATTCAGTGAAGAAGTAG | 4140 |
| Qy | 4294 | aaatttgtaaacatacatattgctgtatttaattccata-----tatt | 4337 |
| Db | 4141 | ATGTGTGTTACAGGTAGCTCTCTGTGCTGATGATATTCACAGGAAGAACAGAGGAGGAGG | 4200 |
| Qy | 4338 | tagccccattcacactaggtcacgcgcgcgaagtgcgtacaaagggaaacgcgcgataacaaatg | 4397 |
| Db | 4201 | AAGTTTCTACAGTAAATGTACGGCGCGAAGTGCTTAAAGGGAGACGGCGATGACAAATG | 4260 |
| Qy | 4398 | taataataactttctctgtgaagacgtttgaaaaatctttttttttttttttttttttttt | 4457 |
| Db | 4261 | TAATTATAACTTCTCTGTGAAGACTTTGGAAAAATC---TTTTTTTTTTTTTTTTTTTTTTT | 4317 |
| Qy | 4458 | ggtaagccttgaggtcgaataaagccttttgatgcacaaaaatgggactgctgaagagtgcga | 4517 |
| Db | 4318 | GGTCAAGCTTGTAGGCTGAAATAAAGCCTTTGATGTGCACAAAATGGCAATGCTGTGAAGAGTGGGA | 4377 |
| Qy | 4518 | cagttggaccttaactttgtgaccccatcacatttgtgtcacatgcttttagccatacac | 4577 |
| Db | 4378 | CAGTTGGACCTTAACCTTTGGTGACCCCATACGTTTGTGGTGCATGCTTTAGCCATACACA | 4437 |
| Qy | 4578 | tggtaacattgactatggagctctgtgaagagttaagtgcgaatggctatgttagacataa | 4637 |
| Db | 4438 | TGGTAAACATTGACTATGGAGCTTGTGTGAAGGTGAATGTGCGCATGGCTATGTAGACATAA | 4497 |
| Qy | 4638 | agaagaacttgtaataacttttttttttttttttttttttttttaagtgttcgtgaagtgcctt | 4697 |
| Db | 4498 | AGAAAGAACTTGTAAATAATCTTTTCTTTTCTTTTAAATGTTCTGATTTCTGAAGTGCCTT | 4557 |
| Qy | 4698 | gtatagcttttactcgcgctttaactgacagctaccgcgacttttattggatctattga | 4757 |
| Db | 4558 | GTATAGCTTTTATCTCGGCTTTAAACTGACAGTACCACGACTGTTTATVTGGATCTATTGA | 4617 |
| Qy | 4758 | tttgaagaagaattgttaggataagatctttaagcagctaactgtccagtgctttgtatttgta | 4817 |

| | | | |
|------------|--|---|-------------|
| Db | 4618 | TTTTGAAAGAAGTAATTTGGTTAGGATAGATCTTTAAGCAGCAATATCTGTACAGTCTTTGTATTTGTA | 4617 |
| Qy | 4818 | ttttctgcaattttactgtgaaaaaaatttggtttcaacaattgggtgtcattttcttga | 4877 |
| Db | 4678 | TTCTCTGCAATTTTACTGTGAAAAAAAATTTGTTTTCAACAAATGGTGTCATTTTCTTGA | 4737 |
| Qy | 4878 | tgtaactatttggtagagagttaaatgggtctcttccctttgtgtatcttacctagtgttt | 4937 |
| Db | 4738 | TGTCACCTATTGTTGGAGAGTTAAATGGTCTCTTCCCCTTTGTGTATCTTTACCTAGTGT | 4797 |
| Qy | 4938 | actctgggcaaccttaattctcaagagtgctaaattgtctgccaattacacagaagat | 4997 |
| Db | 4798 | ACTCCTGGGCACCCCTTTAAATCTTCAGAGGTGCTAAATTTGTCTGCCATTACACAGAAGAT | 4857 |
| Qy | 4998 | gctctgtagagaggacaaccatgcaaatgtgaaatagctcctgaagtctcttgatttact | 5057 |
| Db | 4858 | GCCTCTGATAGGAGGACAACCATGCAAAATGTGGAATAGTCTCTGAAGTCTTTGGATTACT | 4917 |
| Qy | 5058 | ttacacctcagttattgtgtccagaaattttctggcctttcatggcaatgaaaatttt | 5117 |
| Db | 4918 | TTACACTCAGTATTGATTGTTGCCAGAAATTTCTGGCCTTTTCATGGCAATGAAATTTT | 4977 |
| Qy | 5118 | aagaagaagatttaagttattttaaatttaagagtgctgttataaaaaataatgtactgaa | 5177 |
| Db | 4978 | AAGAAGAAAGATTTAAAGTATTTTAAATTTTAAAGAGTGTGTTTATAAAATAATGTACTGAA | 5037 |
| Qy | 5178 | ttctttatcccatttatcatccttctcagtttttatttaactctactgtatcataataaattc | 5237 |
| Db | 5038 | TTCTTTATCCCATTTTATCATCTTTTCAGTCTTTTATTAATCTACTGTATCAATAAAATTC | 5097 |
| Qy | 5238 | tgtaatttggaatg 5250 | |
| Db | 5098 | TGTAATTTGAATG 5110 | |
| RESULT | 4 | | |
| LOCUS | HSU50533 | 1852 bp | 27-NOV-1996 |
| DEFINITION | Human BRCA2 region, mRNA | | |
| ACCESSION | U50533 | CG008 | |
| VERSION | U50533.1 | GI:1531605 | |
| KEYWORDS | human. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| | Couch, F.J., Rommens, J.M., Neuhausen, S.L., Belanger, C., Dumont, M., | | |
| | Kenneth, A., Bell, R., Berry, S., Bogden, R., Cannon-Albright, L., | | |
| | Farrid, L., Frye, C., Hattler, T., Janecki, T., Jiang, P., Kehrer, R., | | |
| | Leblanc, J.-F., McArthur-Morrison, J., McSweeney, D., Miki, Y., | | |
| | Peng, Y., Samson, C., Schroeder, M., Snyder, S.C., Stringfellow, M., | | |
| | Stroup, C., Swedlund, B., Swensen, J., Teng, D., Thakur, S., Tran, T., | | |
| | Tranchant, M., Welver-Feldhaus, J., Wong, A.K.C., Shizuya, H., | | |
| | Labrie, F., Skolnick, M.H., Goldgar, D.E., Kamb, A., Weber, B.L., | | |
| | Tavtigian, S.V. and Simard, J. | | |
| TITLE | Generation of an integrated transcription map of the BRCA2 region | | |
| JOURNAL | on chromosome 13q12-q13 | | |
| MEDLINE | Genomics 36 (1), 86-99 (1996) | | |
| REFERENCE | 9641650 | | |
| AUTHORS | 2 (bases 1 to 1852) | | |
| TITLE | Simard, J. | | |
| JOURNAL | Direct Submission | | |
| | Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular | | |
| | Endocrinology, CHUL Research Center, 2705, Boulevard Laurier, | | |
| | Quebec City, Quebec G1V 4G2, Canada | | |
| FEATURES | Location/Qualifiers | | |
| | 1..1852 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /chromosome="13" | | |
| | /map="13q12-q13" | | |

Db 1321 CAGAACGGATGAAATGCTTTATTTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAG 1380
QY 1537 cattgaatgaatgtgaaatgtcaaaatctgtccgacatcaagtaagagatttgcttg 1596
Db 1381 CATTGAATGAATGTGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTG 1440
QY 1597 acttgattaaagcaacccaaacagatgcccagtgccaagtgccatatatttcaaaagtgtgg 1656
Db 1441 ACTTGATTAAGCAACCCAAACACATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGG 1500
QY 1657 ttatacaagaaatttaacctgatacctgtgtaaggctcaggatttcaatgaagaaattcacac 1716
Db 1501 TTATTACAAGAAATTTACCTGATCTGCTGTAAGGCTCAGGATTTTCATGAAGAAATTCACAC 1560
QY 1717 agtgttagaagatgtagaagaaataagaagacagtttagaagtaactgttagtccaacat 1776
Db 1561 AGGTGTTAGAAGATGATGAGAAATAGAAAGCAGTTAGAAAGTACTGTTTAGTCCAACAT 1620
QY 1777 gctcctgcaagcaggctgaaggttgtgtgctgaaataaactaaagaagtgggcaacccca 1836
Db 1621 GCTCCTGCAACGAGCTGAAGGTTGTGTGCGTGAATAACTTAAGAAGTTGGGCAACCCCA 1680
QY 1837 aacgcctacaaatcccttccctggaatgatcaagtttctctcttgagaggatgacaactg 1896
Db 1681 AACAGCCTTACAAATCCTTTCTCTGAAATGATCAAGTTTCTCTTGAGAGGATAGCACCTG 1740
QY 1897 tgcacatagatacgaatctatacagttgctcttatttaaacaaagtgaacaaatcaatagatg 1956
Db 1741 TGCACATAGATACCGAATCTATCAGTGTCTTTATTAACAAAGTAGCAAAATCAATAGATG 1800
QY 1957 gaacagcagatgataagaatgaggtgttccaactgatcaagccatcagagcaggtcttg 2016
Db 1801 GAACAGCAGATGATGAAGATGAGGTTGTCCAACTGATCAAGCCATCAGACGAGTCTTG 1860
QY 2017 aactgcttaaggtaactctcattacacatcccatctcattctctgtgtaaacatttg 2076
Db 1861 AACTGCTTAAGGTACTCTCAATTTACATCCATCCATCTCATTTCTGCTGAACATTTG 1920
QY 2077 aatcattactgctgtctgaaatggatgatgaaagtagcagaagctgcactacaaa 2136
Db 1921 AATCATTAAGGTGCTGTGAAATGGATGATGAAAGTAGCAAGAGCTGCATACAAA 1980
QY 2137 ttctcaaaaaacacaggaagcaaaattgaagaggattttccacacatcagatcagcccttg 2196
Db 1981 TTTTCARAAACACAGGAAGCAAAATGGAAGAGGATTTCCACACATCAGATCAGCCCTGC 2040
QY 2197 ttcctgtttttacatacaaaaactcaaaaaagaccccccccgtaagccaaaatatgccattc 2256
Db 2041 TTCCTGTTTTACATCACAAATCTAAAAAGGACCCCGCTCAAGCCCAAAATATGCCATT 2100
QY 2257 attgtatccatgcgatatatttctgtaaaagagacccagtttgacagatatattgagccctc 2316
Db 2101 ATTGTATCCATGCGGATATTTTCTAGTAAGAGAGCCCGAGTTTGCACAGATATTTGAGCCTC 2160
QY 2317 tgcataagcgtctagatccagcaacctggaaactctcatacaaccattggttactattg 2376
Db 2161 TGCATAAGAGCCTAGATCCCAAGCAACCTGGAAACATCTCATAAACCATTTGTTACTATTG 2220
QY 2377 gtcataatgtcctctgcaactgatacaattgtgtcctcttgaaatctctgggtagcta 2436
Db 2221 GTCATATTTGCTCTCTTGACCTGATCAATTTTGCTGTCTCTTTGAAATCTTTTGGTAGCTA 2280
QY 2437 ctttcattgtgaagatcttctaatgaatgatcggtcttccagggaaaaaagacaactaaac 2496
Db 2281 CTTTCATTGTGAAAGATCTCTCATGTAATGATCGGCTTCCAGGGAAAAAGACAACCTAAAC 2340
QY 2497 ttgtgggtccagatgaagaagatctcctctgagacaaatgggtcaaaattcaggtctattaaa 2556
Db 2341 TTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAAATGGTCAAAATTCAGGCTATTAAAA 2400
QY 2557 tgaatgttcgactgtggaatgaaaaataataacagtaaatcaagtaaatcgaactctacct 2616
Db 2401 TGATGGTTCGATGGCTACTTTGGAATGAAAAATAATACAGTAAATTCAGGAACCTTCTACCT 2460

QY 2617 taagcttgctaaacaataatttgcatagtgatgagacttgacagacaggggaaaaatta 2676
Db 2461 TAAGATTGCTTAACAAACAATATTTGCATAGTGATGAGACTTGCACAGAACAGGGGAAAAATTA 2520
QY 2677 gtaaacacagatattgcacgtctgagacttgctgctggagtgctattgtgaagctggcac 2736
Db 2521 GTAAACCAGATATGTCACGCTGAGACTTGTCTGGGAGTGCATTGTGAAGCTGGCAC 2580
QY 2737 aagaacctgttaccatgaataatcatcaatttagaacaataatcagctatgtgcattagcta 2796
Db 2581 AAGAACCTGTTTACCATGAATCATCATATAGAACAAATATCAGCTATGTGCATTAGCTA 2640
QY 2797 tcaacgatgaatgctatacaagtaagaagttgttgcacagaaacttcaaaaaagccttt 2856
Db 2641 TCAACGATGAATGCTATCAAGTAAGCAAGTGTGTGGCCAGAAACTTTCACAAAGGCCCTTT 2700
QY 2857 cccgtttacggcttccacttgagtatatggcaatctgtgcccctttgtgcaaaaagactctg 2916
Db 2701 CCCGTTTACGGCTTCCACTTGGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTG 2760
QY 2917 taaaggaggaagagctcaatgtctaggccaattcttgggtgaaaaataataaatgaaggcggg 2976
Db 2761 TAAAGGAGAGAAAGAGCTCATGCTAGGCAATGTTTGGTGAATAATATAAATGTAAGCGGG 2820
QY 2977 agtatctgaagcagcagcatgcaactgttagtgaaaaattattgtctctctaccagagtatg 3036
Db 2821 AGTATCTGAAGCAGCATGTCAGCTGTTAGTGAATAATATTGTCTCTTACACAGAGTATG 2880
QY 3037 ttgttccatataacaaatccacttttggcacatgaccagattatgtcaaaagtacaggata 3096
Db 2881 TTGTTCCATATACAATTCACCTTTTGGCACATGACCAGATATATGTCAAACTACAGGATA 2940
QY 3097 ttgaacaaacttaagatggttaagaaatgtcttgggtttgttctggaataataatgggcta 3156
Db 2941 TTGAACAACCTTAAAGATGTTTAAAGAAATGCTTTGGTTTGTCTGCTGGAATAATTAATGGCTA 3000
QY 3157 aaaaataaaacagtcacgcttttcaagaagatggtagaataattataaacaacaa 3216
Db 3001 AAAATGAAAAATACAGATCACCGCTTTTATCAGAAAGATGGTAGAAAAATATTAAACAAACAA 3060
QY 3217 aagatgcccaggaccagatgatcaaaaatgaatgaaaaaactgtacactgtgtgtgatg 3276
Db 3061 AAGATGCCAAAGGACCAGATGATGCAAAAATGAATGAAAAAACTGTACACTGTGTGTGATG 3120
QY 3277 ttgcatgaatatcatcatgtcaaaagagtaatacacagtttgggaatctcctctaaagacc 3336
Db 3121 TTGCCATGAATATCATCATGTCAAAAGAGTACTACATACAGTTTGGAAATCTCTCTAAAGACC 3180
QY 3337 cggtaactaccagctcggttcttctcaactcaacctgacaagaatttcagtaaacacccaaaaatt 3396
Db 3181 CGGTACTACCAGCTCGTGTCTTCACTCAACCTGACAAGAAATTTTCAGTAACACCAAAAAAT 3240
QY 3397 atctgctctctgaaatgaaatcatttttcaactctggaacacccctaaacacccaatgttc 3456
Db 3241 ATCTGCTCTCTGAAATGAAATCAATTTTTCACCTCTGGAACAACTTAAACACCAATGTTC 3300
QY 3457 tagagctgtttaacagccacttctcaatgaaggaagcaatctcaagaccacaaatcaatcaac 3516
Db 3301 TAGAGCTGTTTAAAGAGCCACTTTTCATCAGCAGGCAAGCAATCTCAGACCAAAATCATCAC 3360
QY 3517 gaatgaaactgttaagaataatcaagcagcagctcaaatccaaagctctcctggaagaataa 3576
Db 3361 GAATGGAAACTGTGAAGCAATCAAGCAGAGCTCAAAATCCAAGCTCTCTCTGGAAGATAA 3420
QY 3577 aggggagcttgatgttctgaaatggatcacagtgaaatgaagatttacacaatgtctt 3636
Db 3421 AGGGAGGCTTGTAGTTCTGAATGGATCACAGTGAATAATGAAGATTACACAATGTCTT 3480
QY 3637 caccttgcgggggaaaaaagtgtaagaagagacagctctgatcttgaaggtctgaat 3696
Db 3481 CACCTTTGCCGGGAAAAAAGTACAAAGAGAGACGACTCTGTATCTTTGTAAGGTCTGAAT 3540

|||||
3482 TTTTCACTCTCTGAAAAACCTAAACAACTATGTTCTAGAGCTGTTAAACAGGCACATTT 3541
QY |||||
3481 catcagcaggcaagaatctcagaccacaatcatcagcaatggaactgtaagcaatgcaaa 3540
Db |||||
3542 CATCAGCAGGCAAGCAATCTCAGCCAAATCATCAGCAATGGAACACTGTAACATGCA 3601
QY |||||
3541 gcagcagctcaaatccaaagctctccctgggaataaaaggggggcttgatgtctgaaa 3600
Db |||||
3602 GCAGCAGCTCAATCCAAAGCTCTCTGGGAATAAAGGGAGGCTTGATAGTTCTGAAA 3661
QY |||||
3601 tggatcacagtgaataaagaattacacaaatgtcttcacactttgcgggaaaaaaagtg 3660
Db |||||
3662 TGGATCACAGTGAATAATGAAGATTAACAAATGCTTCACCTTTGCCGGGAAAAAAGTG 3721
QY |||||
3661 acaagagagcagctctgtatctgttaaggtctgaaatggagaagcctagagcaggaaaa 3720
Db |||||
3722 ACAAGAGAGAGACTCTGATCTTGTAAAGTCTGAATGGAGAAGCCTAGAGCGCAGAAAA 3781
QY |||||
3721 aaacgcgcgtcacagaaagagggagaaataggatgagtagtgaactgaactaagtgtac 3780
Db |||||
3782 AAACGCCGTCACAGAACAGAGAGAGAAATAGGTATGGATGACTTACTTAAGTTGGTAC 3841
QY |||||
3781 aggaacagaacacctaaaggcagtcagcgaagtgcgaaaagagccatacagcgtctcagaat 3840
Db |||||
3842 AGGAACAGAAACCTTAAAGGCAGTCAGCGAAGTCGGAAGAGAGGCCATACGCGTTCAGAA 3901
QY |||||
3841 ctgtagaacagcagctggcctgaggaagagggctcaagaagatatattagaagaatgaag 3900
Db |||||
3902 CTGATGAACAGCAGTGGCTTGAGGAAGAGGCTCAAGAAGATATATTAGAAAATGAAG 3961
QY |||||
3901 atgaacagaatagtcgcgcaaaagggtaaaagagccagcaacaaacccctctgtg 3960
Db |||||
3962 ATGAACAGATAGTCCGCAAAAGGGTAAAGAGGCGCCAGCCACCAAAACCTCTTG 4021
QY |||||
3961 gaggctacacaaagagagcacaatgaaactctcaaaaggaagcaaaacaaat 4020
Db |||||
4022 GAGGTACACCAAAAGAGAGCCCACTGAATACTCTTAAAAAGGAAGCAAAAAAAT 4081
QY |||||
4021 ctggacctccagcaccagagaggagggaagaagaagaagcaagtggaataacggaac 4080
Db |||||
4082 CTGGACCTCCAGCACAGAGGAGGAGGAAGAAGAGAAAGCAAAAGTGGAAATACGGAAC 4141
QY |||||
4081 aqaagtccaaagcaaacagaccagctgtcaaggagagcagcagagagcagaaatctc 4140
Db |||||
4142 AGAAGTCCAAAGCAACAGCACCGACCTGTCNAGGAGAGCAGCAGAGAGCAAAATCTC 4201
QY |||||
4141 ctgaactagtgaattgaatccacacagtcacacaccagaaagagcagaggaagaccat 4200
Db |||||
4202 CTGAATCTAGTGAATTCACACAGTCCACACACCACAGAAAGGAGGAGAGACCAT 4261
QY |||||
4201 caaaacgcgcac 4260
Db |||||
4262 CAAAGACGCATACCATCACACCAACCAAAAGAAATGTCGCTGAGGAGCGCTCCCAACAG 4321
QY |||||
4261 aa---accatttcaaatatttttgcgaaggtctctaaatttgtaaacatacatattgct 4317
Db |||||
4322 CAGCTACTAGGAAATGATTCAAGTGAAGAGTAGATGTGTTTTCAGGGTAGCTCTCCTG 4381
QY |||||
4318 gtatttaattccatat-----attagcccccattacactaggtacgcg 4361
Db |||||
4382 TCGATGATATTCACAGGAGAAACAGAGAGGAGGAAGTTCACAGTAATATGACGGC 4441
QY |||||
4362 ggcgaagtctaaagggagcgcatgacaaatgtaatttaacttctctctgtgaaa 4421
Db |||||
4442 GGGGAAGTGTAAAGGGAACGCGATGACAAATGTAATTAATCTTCTCTGTGAAA 4501
QY |||||
4422 gcttggaataatcttt 4481
Db |||||
4502 GCTTTGAAAAATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTCAAGCTTGAGGCTGAATAAG 4561
QY |||||
4482 cctttgatgacaaaatggagctgctgaagagtggaagcttgacaccttctgtgacc 4541
|||||

Db 4562 CCTTTGATGCACAAAAATGGGACTGCTGAAGAGTGGACAGTGGACCTTACTTTGGTGACC 4621
QY |||||
4542 ccatcacatttggtgcacatgcttttaagccatacacacatggaacattgactatgagctctt 4601
Db |||||
4622 CCATACATTTTGTGTGCACATGCTTTAGCCATACACATGGAACATTTACTATGGAGTCTT 4681
QY |||||
4602 gtgaaagtgtaatgtgagatggtctatgtagacataaaagaagaacttgtaaatatctttt 4661
Db |||||
4682 GTGAAGTGTAAATGTGCGATGCTATGTAGACATAAAGAGAAACTTGTAAATATCTTTT 4741
QY |||||
4662 ttcttttttaagtctcttcatttcgaagtgtgttatagcttttatctgcgagcttta 4721
Db |||||
4742 TTCTTTTTTTTAAATGTTTCTGATTTCTGAAGTGTGTATAGCTTTTAICTGCGGCTTTA 4801
QY |||||
4722 aactgaacagctaccgagctgtttattggatctctattgattgaaagaatttcttaggtag 4781
Db |||||
4802 AACTGACAGTACCCGACCTGTTTATTTGGATCTATTGATTTGAAAGAAATTTGTTAGGATAG 4861
QY |||||
4782 atcttaagcagtaactgtcagtggtttgtattgtatttcatttcgaattttactgtgaaaa 4841
Db |||||
4862 ATCTTAAGCAGTAATCTGTCACTGTTTGTATTGTATTCTCTGCAATTTTACTGTGAAAA 4921
QY |||||
4842 aaaaattgttttcaacaattggtgtcattttcttgatgtcactatttctgtgagaggttaa 4901
Db |||||
4922 AAAATTTGTTTCAACAATFTGTGTCTATTTCTTCTGATGTCATATTGTTGAGAGATTAA 4981
QY |||||
4902 atggctctctccctttgtgtatcttaccctagttgttactcctgggcaaccttaacttcaa 4961
Db |||||
4982 ATGGTCTCTCCCTTTGTCTATCTTACCTAGTGTGTTTACTCTCTGGCAGCCCTTAATCTTCA 5041
QY |||||
4962 gaggtgctaaaattgtcgcattacaccagaaggatgcctctgatagagagacaaccatg 5021
Db |||||
5042 GAGGTGCTAAATTTGTGTCATTTACACCAGAAGGATGCCTCTGATAGGAGGACAAACCATG 5101
QY |||||
5022 caaattgtaaatagctcctgaggtcttgattacttaccctcagttattgattgttctcc 5081
Db |||||
5102 CAAATTTGTGAATATGTCCTGAAAGTCTTGGATTACTTTACCTCAGTATTGATTGTCTCC 5161
QY |||||
5082 cagaattttctggcctttcagcaatgaaattttaagaagaagaatttaagaattttt 5141
Db |||||
5162 CAGAAATTTCTGGCCTTTCATGCAATGAAATTTTAAAGAAGAAATTTAAAGATATTTT 5221
QY |||||
5142 aattttaagagtgattataaaataatgactgactgaattctttatccattttatcatcct 5201
Db |||||
5222 AATTTTAAAGAGTGTGTTATATAAATAATGACTGTAATCTTTTATCCATTTTATCATCCT 5281
QY |||||
5202 ttcagttttttatcaatctactgtatcaataaaattctctgtaatttgatgaagt 5253
Db |||||
5282 TTCAGTTTTTATTAACTACTGTATCAATAAATAATCTGTAAATTTGAATGAGT 5333

RESULT 3
AB023196 5110 bp mRNA PRI 16-JUN-1999
LOCUS Homo sapiens mRNA for KIAA0979 protein, partial cds.
ACCESSION AB023196
VERSION AB023196.1 GI:4589601
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_lib:pbluescriptII
SK plus clone:hj07056.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
xiii. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 6 (1), 63-70 (1999)
JOURNAL 9246063
MEDLINE
REFERENCE 2 (bases 1 to 5110)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.

| | | | |
|----|------|---|------|
| Qy | 2341 | acctggaaacatctcataaacacacattgggttactatttggtgcataattgctctctctctgcacctg | 2400 |
| Db | 2402 | ACCTGGAAACATCTCATTAACACCATTTGGTTACTATTATTTGGTCATATTGCTTCCTCTGACACTG | 2461 |
| Qy | 2401 | atcaatttgcctctcttggaaatcttgggtagctacttctcatigtgaaagatcttctca | 2460 |
| Db | 2462 | ATCAATTTTGCTGCTCTTTTGAAATCTTTGGTAGCTACTTTTCATTTGTGAAAGATCTTCTCA | 2521 |
| Qy | 2461 | tgaatgcctggctccagggaagaaacactaaacttgggtttccagatgaagaagtat | 2520 |
| Db | 2522 | TGAATGATCGGCTTCCAGGGAAAGAACACAACTTAAACTTTTGGGTTCCAGATGAAGAAGTAT | 2581 |
| Qy | 2521 | ctctcagacaatgggtcaaaattcaggctattaaaatgatggttcgatggctactctggaa | 2580 |
| Db | 2582 | CTCTGAGACAATGGGTCAAAATTTACGGCTATTAAATGATGGTTCGATGGCTACTTGGAA | 2641 |
| Qy | 2581 | tgaataataatcacagttaaactcaggaaactctaccttaagatttgctaaacaacataatgc | 2640 |
| Db | 2642 | TGAAAAATAATCACAGATTAATCAGGAACCTTACCTTAAAGATTGCTAAACAACAATATTGC | 2701 |
| Qy | 2641 | atagctgatgagacttgacagaacacaggggaaataattagtaaacacagatatgtcacgctctga | 2700 |
| Db | 2702 | ATAGTGTAGGAGACTTTGACAGAACAGGGGAAATAATTAGTAAACACGATATGTACAGCTCTGA | 2761 |
| Qy | 2701 | gacttgcctgggagtgctattgtgaagctggcacagaacccctgttaccatgaaataca | 2760 |
| Db | 2762 | GACTTGTCTGTGGAGTGCTATTGTTGAGCTGGCACAGAACCCTGTTACCATGAATCA | 2821 |
| Qy | 2761 | tcacattagaacaatacagctatgtgctattagctatcaacgatgaatgctatacaagtaa | 2820 |
| Db | 2822 | TCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAA | 2881 |
| Qy | 2821 | gacaagtgtttgcccagaacttcaaaagcccttcccggtttacgcttccacttgagt | 2880 |
| Db | 2882 | GACAAAGTGTTTGGCCCAAGAACTTCAACAAGCCCTTCCCGTTTACGGCTTCCACTTGA | 2941 |
| Qy | 2881 | ataatgccaatctgtgccctttgtgcaaaagactcctgaaaggagagaagctcatgcta | 2940 |
| Db | 2942 | ATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTTAAAGGAGAGAAGAGCTCATGCTA | 3001 |
| Qy | 2941 | ggcaatgtttggtagaaaataataatgtaggcgagatctctgaagcagcatgcagctg | 3000 |
| Db | 3002 | GGCAATGTTTGGTGAAAAATATAAATGTAAAGCGGAGTATCTGAAGCAGCATGACAGCTG | 3061 |
| Qy | 3001 | ttagtgaataattatgtctctctacagagtatgtgttccataatacaattcaccttt | 3060 |
| Db | 3062 | TTAGTGAAAAATTATGTCTCTCTACAGAGTATGTTGTTTCCATATACAAATTCACCTTT | 3121 |
| Qy | 3061 | tggcacatgacccagattatgcaaaatcacagatatgaacaacttaagatgtaaaag | 3120 |
| Db | 3122 | TGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACCTTAAAGATGTTAAG | 3181 |
| Qy | 3121 | agatgcttctgttcttggaataataatggcctaaaaatgaaataacagtcacgctt | 3180 |
| Db | 3182 | AATGCTTTTGGTTTCTCGAAATATTAAATGGCTAAAAATGAATAACACGTACAGCTT | 3241 |
| Qy | 3181 | ttatcagaagaatgtagaaaataattaacaacaagaatgccccagacagatgatg | 3240 |
| Db | 3242 | TTATCAGAAAAGATGTAGAAAAATATTAACAAACAAAGAATGCCCAAGGCCAGATGATG | 3301 |
| Qy | 3241 | caaaaatgaatgaaaaaactgtacactgtgtgatgttgccatgaatatcatcatgtcaa | 3300 |
| Db | 3302 | CAAAAATGAATGAAAAATGTACTACTGTGTGTATGTTGCCATGAATATCATCATGTCAA | 3361 |
| Qy | 3301 | agagtactacatacacagtttggaatctctcctaaagaccgggtactaccagctcggtttctca | 3360 |
| Db | 3362 | AGAGTACTACATACAGTTTGGAAATCTCCTTAAGACCCGGTACTACCAGCTCGGTTTCTCA | 3421 |
| Qy | 3361 | ctcaacctgacaagaatttcagtaacacccaaaaatattctgcctcctgaaatgaatacat | 3420 |
| Db | 3422 | CTCAACCTGACGAAGAATTTCAGTACACAAAAAATTATCTGCCCTCTGAAATGAATATCAT | 3481 |
| Qy | 3421 | ttttactcctcagaaacccataaacacccaagtgtctlagagctgttaaacaagccacttt | 3480 |


```

/number=33
4255..4307
/note="matches EST T64867 from clone 22264"
exon
4299..4434
/number=34
join(4398..4533,4518..4660,4659..4675,4674..4693)
/note="matches EST HA4781 from clone 188032"
exon
4435..4744
/number=35
4537..5022
/note="matches EST AI681422 from clone IMAGE:2272632"
misc_feature
4537..4892
/note="matches EST AA911125 from clone IMAGE:1416862"
misc_feature
4538..4969
/note="matches EST AI275866 from clone IMAGE:1877482"
misc_feature
4538..4947
/note="matches EST AA987361 from clone IMAGE:1603350"
misc_feature
4538..4943
/note="matches EST AI299646 from clone IMAGE:1898010"
misc_feature
4562..5003
/note="matches EST AA447433 from clone 784555"
misc_feature
4571..4854
/note="matches EST H93424 from clone 220700"
misc_feature
join(4573..4649,4650..4815)
/note="matches EST W20070 from clone 306049"
misc_feature
complement(4629..4937)
/note="matches EST AW089775 from clone IMAGE:2593146"
misc_feature
4746..4994
/note="matches EST AA747568 from clone IMAGE:1269696"
misc_feature
complement(4803..5332)
/note="matches EST AI911784 from clone IMAGE:2329873"
misc_feature
complement(4828..5296)
/note="matches EST AA563884 from clone IMAGE:1019990"
misc_feature
4861..5243
/note="matches EST D20082 from clone pm1283."
misc_feature
complement(join(4872..4921,5281..5332))
/note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature
join(4872..4921,5281..5371)
/note="matches EST AA177347 from clone 621810"
misc_feature
complement(4872..4921)
/note="matches EST AW048929 from clone
UT-M-BH1-amo-c-09-0-UI"
4872..4921
/note="matches EST AA154932 from clone 599194"
misc_feature
complement(4876..5333)
/note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature
complement(join(4922..4975,5281..5332))
/note="matches EST A1012649 from clone RPLBA02"
misc_feature
complement(4922..4975)
/note="matches EST AA899586 from clone
UT-R-E0-cy-e-02-0-UI; matches EST AI72172 from clone
RMUBV82"
misc_feature
complement(4931..5332)
/note="matches EST AA625961 from clone 745478"
misc_feature
complement(4951..5332)
/note="matches EST A1140790 from clone IMAGE:1566222"
misc_feature
complement(5021..5309)
/note="matches EST H45693 from clone 188292"

Query Match          96.5%; Score 5084.4; DB 92; Length 7444;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 5172; Conservative 0; Mismatches 81; Indels 19; Gaps

QY      1  ccggagagcccgagtgagcgagtagcgatcggaacccggagggtagaatatatt 60
      |||
Db       62  CCGGAGAGCCCGGAGTGAGCGGAGTAGCGAGTCGGCAACCGGAGGGGTAGAAATATT 121
      |||

QY      61  ctgtcatgctcattcaagaactagacacatgattggaataattacatcccgctggg 120
      |||

Db       122  CTGTCTATGGCTTCATTCAAGACACTAGGACCAATGATGGAAAAATTACATATCCGCTGGGG 181
      |||

QY      121  tcaaggaataatcagataaaatcttaagagagagatggtgagacgattaaagatggtt 180
      |||

```

| | | | |
|----|------|--|------|
| Db | 182 | TCAGGAAATATCAGATAAAATATCTAAAGAGGAGATGCTGACAGATTAAAGATGGTTG | 241 |
| Qy | 181 | tgaaaacttttattgatatgaccagactctgaagaagaagaagagctttatttaaacc | 240 |
| Db | 242 | TGAAAACTTTATGGATATGGACGAGACTCTGAGAGAGAAAGGAGCTTTATTTAAACC | 301 |
| Qy | 241 | tagctttacatcttgccttcagattttttctcaagcatcctgtaaagatgttcgcttac | 300 |
| Db | 302 | TAGCTTTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTAC | 361 |
| Qy | 301 | tggtagcctgcctgcctgcctgatatatttcagagatttatgctcctgaagctccttacacat | 360 |
| Db | 362 | TGGTAGCCTGCTGCTTGCATATTTTTCAGGATTTAIGCTCTGAAGCTCCTTACACAT | 421 |
| Qy | 361 | ccctgataaactaaagatatattatgtttataacaagacagttgaaggggctagagg | 420 |
| Db | 422 | CCCCTGATTAACCTAAAGATATATTTATGTTTATAACAAGACAGTTGAAGGGCTAGAGG | 481 |
| Qy | 421 | atacaaagagcccaactcaatagggtatttttatttacttgagaacatgcttgaggta | 480 |
| Db | 482 | ATPACAAAGAGGCCACAAATTCAAATAGGTATTTTTTATTACTTGAGAACATTTGCTTGGGTCA | 541 |
| Qy | 481 | agtcataaacatatgctttgagttagaagatagcaatgaatttcaccagctataca | 540 |
| Db | 542 | AGTCATATAACATATGCTTTCAGTTAGAAGATAGCAATGAAATTTTCCACGAGTATACA | 601 |
| Qy | 541 | gaaccttatatttcagdtataaacaatgccacaatcagaagaatccatatgcacatgtag | 600 |
| Db | 602 | GAACCTTATTTTCAGTTATAACAAATGCCACAATCAGAAAGTCCATATGCATCGGTAG | 661 |
| Qy | 601 | accttatagctctattattgtgaagtgatcacagtgctcagagagcttttgatcacgg | 660 |
| Db | 662 | ACCTTATAGACTCTATATTTTGTGAAGTGATACAGTGTCTCAGGAGCTTTTGGATACGG | 721 |
| Qy | 661 | ttttagtaaatcgtgtacctgctatagaatttaaaacaagcaagcatatgattggcaa | 720 |
| Db | 722 | TTTTTAGTAAATCTGCTTACCTGCTCATAGAATTTAAACAGCAGCATATGATTTGGCAA | 781 |
| Qy | 721 | agcctttactgaagagcagctcaagctattgagccatatattaccacttttttaatc | 780 |
| Db | 782 | AGGCTTTACTTGAGAGCAGCTCAAGCTATTAGCCATATATTACCAATTTTTTTAATC | 841 |
| Qy | 781 | aggtctctgatgttgggaaaaacatctatcaggagttgtcagagcatgtcttgacttaa | 840 |
| Db | 842 | AGGTTCTGATGCTTTGGGAAAACTCTATCAGCGATTTGTCAGAGCATGTCTTTTGACTTAA | 901 |
| Qy | 841 | ttttgagctctacaatattgatgctattgtcgtcctctgtttaccocagcttgaat | 900 |
| Db | 902 | TTTTTGGAGCTCTACAATTTGATAGTCATTGCTGCTCTCTGTTTACCCAGCTTGAAT | 961 |
| Qy | 901 | ttaaatataagagcaatgataatgaggagcgcctacaagtgtttaaactactggcaaaa | 960 |
| Db | 962 | TTAAATTTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTGTGTTAACTTAC'TGGCAAAA | 1021 |
| Qy | 961 | tgtttgggcaagagattcagaattggctctctcaaaacaagccaactttggcgagtctact | 1020 |
| Db | 1022 | TGTTTGGGGCAAGGATTCAGAAATGGCTTCTCAAAACAGCCACTTTGGCAGTGTACT | 1081 |
| Qy | 1021 | tgggcaggttttaatgatatocattgataccaatccgcctggaatgtgtgaatttgcctagcc | 1080 |
| Db | 1082 | TGGCAGGTTTTAATGATATCATGTACCAATCGCGCTGGAATGTGTAAATTTGCTAGCG | 1141 |
| Qy | 1081 | attgtctcatgaaccatccgtatttagcaaaaagacttaacagagtatctttaaagtgaggt | 1140 |
| Db | 1142 | ATTGFTCATGAACCAATCCTGATTTAGCAAAAGACTTTACAGAGTATCTTTAAAGTCAGGT | 1201 |
| Qy | 1141 | cacatgaccttgaggaagctattagacatgatgttatgtgtcaatagttacaagctgcta | 1200 |
| Db | 1202 | CACATGACCCCTGAGGAAGCTATTAGACATGATGTTATTTGTCAATAGTTTACAGCTGCTA | 1261 |
| Qy | 1201 | aaaaagatatcttctggtcaatgatcacttacttaatttttggagagagagacaattag | 1260 |
| Db | 1262 | AAAAAGGATATCTCTGTGCTCAATGATCACTTACTTAAATTTTGTGAGAGACAGAACATTAG | 1321 |

exon

| | | | | |
|------|---------------|--|--|------|
| 4141 | CTGAATCTAGTGC | AAATTGAATCCACAGTCCACACCCACAGAAAGCAGGAGAACCAT | 4200 | |
| Db | | | | |
| | 4201 | caaaaagccatcaccatcaacacaaa | aaatgtgaagtgttaaatattacattc | 4260 |
| Qy | | | | |
| Db | 4201 | CAAAAAGCCATCACCATCAACACCA | AAAAAATGTGAAGTGTAAATATATAGATTTC | 4260 |
| | 4261 | aaaccaattcaaatattatttg | caaaaagttctaaattgttaaacatcacatattgctga | 4320 |
| Qy | | | | |
| Db | 4261 | AAACCAATTTCAAAATATTTTGC | AAAAAGTTCCTAAATTTGTAAACATACATATTTGCTGTA | 4320 |
| | 4321 | tttaaatccatatatttagcccatcacact | agtgacgcggaagtgtctaaaagga | 4380 |
| Qy | | | | |
| Db | 4321 | TTTAAATTCATATATTAGCCCATACACT | AGGTAGGCGGCGAAGTGCTAAAAGGA | 4380 |
| | 4381 | acggcgatgaacaaatgtaattaa | actttctctgtgaagctttggaaaaattctttt | 4440 |
| Qy | | | | |
| Db | 4381 | ACGGCGATGAACAAATGTAAATTA | ATACTTCTCTGTGAAGCTTTGGAAAAATCTTTT | 4440 |
| | 4441 | tttttttttttttttttg | tcaagcttgagctgaaataagacctttgatgcacaaatgg | 4500 |
| Qy | | | | |
| Db | 4441 | TTTTTTTTTTTTTTTTTGGTCA | GCTTGAGGCTGAATAAAGCCTTTGATGCACAAAATGG | 4500 |
| | 4501 | gactgctgaagatggaacagttg | gaccttactttggtgacccacacatttgggtcac | 4560 |
| Qy | | | | |
| Db | 4501 | GACTGCTGAAGAGTGGACAGTTGG | ACCTTACTTTGGTGACCCATACATTTGTGGTCACA | 4560 |
| | 4561 | tgctttagccatacacatggt | taacatgactatgagctctgtgaaagtgtaatgtgca | 4620 |
| Qy | | | | |
| Db | 4561 | TGCTTTAGCCATACACATGGTAA | CAACTGACACTATGGAGCTTTGTCAAAGTGAATGTGCGA | 4620 |
| | 4621 | tgctatgtagacataaagaaga | aacttgtaatatcttttcttttttaattgttctc | 4680 |
| Qy | | | | |
| Db | 4621 | TGGCTATGTAGACATAAAGA | AAACTTGTAAATATCTTTTCTTTTTTTTAAATGTTTC | 4680 |
| | 4681 | tgatttctgaagtcttatagct | ttttatctcgcgctttaaacctgacacgacccgactg | 4740 |
| Qy | | | | |
| Db | 4681 | TGAATTTCTGAAGTGTGTATAG | CTTTTATCTCGCGCTTTTAACTGACAGTACCCGACTG | 4740 |
| | 4741 | tttattggatctattgattt | gaaaagaatttgtagatagatcttaagcagtaattctgt | 4800 |
| Qy | | | | |
| Db | 4741 | TTTATTGGATCTATTGA | TTTGAAGAAGATTTGTAGATAGATCTTAAGCAGTAATCTGT | 4800 |
| | 4801 | cagtggttattgtatttct | gcaatttctacgtgaaaaaatttgtttcaacaat | 4860 |
| Qy | | | | |
| Db | 4801 | CAGTGTGTGTATTTGTATTTCT | CGAATTTTACTGTGAAAAAAATTTGTTTCAACAAT | 4860 |
| | 4861 | tggtgcatttcttgatgcact | attgttggagagtaaattggctctccctttg | 4920 |
| Qy | | | | |
| Db | 4861 | TGGTGTCAATTTCTTGATG | TCACTATTTGTGGAGAGTTAAATGGTCTCTTCCTTTGTG | 4920 |
| | 4921 | tatcttacctagtggtt | taactctgggaccccttaacttcagagtgctaaattgctc | 4980 |
| Qy | | | | |
| Db | 4921 | TATCTTACCTAGTGTTTACTCT | CGGACCCCTTAATCTTCAGAGGTGCTTAATTTGCTCG | 4980 |
| | 4981 | cattacaccagaagatgcctct | gtatagagagcaacocatgcaaatgtgaaatagtcct | 5040 |
| Qy | | | | |
| Db | 4981 | CATTACACCAGAAGATG | CCCTCTGTAGGAGGACAACCATGCAAAATGTGAAATAGTCCT | 5040 |
| | 5041 | gaagtcttggaatacttt | acacctgaattgtgtccacgaatttctgccttc | 5100 |
| Qy | | | | |
| Db | 5041 | GAAGTCTTCGGATTA | CTTTTACACCTCAGTATTTGATTTTGTCCAGAAATTTCTTGCCCTTTC | 5100 |
| | 5101 | atggcaatgaaatttaagaaga | agaatttaaaagtatttaatttaaagagtggtta | 5160 |
| Qy | | | | |
| Db | 5101 | ATGCCAATGAAATTTTAA | GAGAAAGATTTAAAGTATTTTAAATTTTAAAGAGTGTGTTA | 5160 |
| | 5161 | taaaataatgtactgaatt | totttatcccaattttatcactctttcaagtttttataacta | 5220 |
| Qy | | | | |
| Db | 5161 | TAAAAATATGTACTGA | ATTTTATCCCAATTTTATCATCTCTTCAGTTTATTATTAATCTA | 5220 |
| | 5221 | ctgtatcaataaaattctg | taattgtgaatgagt | 5253 |
| Qy | | | | |
| Db | 5221 | CTGTATCAATAAATTC | TGTAAATTTGAATGAGT | 5253 |

QY 901 ttaaaattaaagcaaatgataatgaggagcgccctacaaagtgtttaaactactggycaaaa 960
Db 901 TTAATTTAAAGAGCAATGATAATGAGGAGCGCTCAAGTGTGTTAAACTACTTGGCGAAAA 960
QY 961 tgtttggggcaagattcagaaattggctctcaaaaacagccactttggcagtgctact 1020
Db 961 TGTTTGGGCAAGGATTGAGAAATGGCTTCTCAAAACAGCCACTTTGGCAGTGCTACT 1020
QY 1021 tgggcaaggtttaatgatatccatgctaccaatccgcctggaattgtgaaatttgcagcc 1080
Db 1021 TGGGCAAGGTTTAATGATATCCATGATGACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCC 1080
QY 1081 atgttctcatgaaccatctgatttagcaaaagacctaacagagatatottaaagtgaagt 1140
Db 1081 ATTGTCTCATGAACCATCTGATTTAGCAAAAGACTTAAACAGAGTATCTTAAAGTGAGGT 1140
QY 1141 cacatgacctgaggaagctattagacatgatgttattgtgtcgaatgattcacagctgcta 1200
Db 1141 CACATGACCTCGAGGAAGCTATTAGACATGATGTTATTGTCTCAATAGTTACAGCTGCTA 1200
QY 1201 aaaaagatatcttctgtgcaatgatcaactacttaatttbtgtgagagagagaacattag 1260
Db 1201 AAAGGATATTTCTGTGTCAATGATCACTTACTTAATTTTGTGGAGAGAGAACAATTAG 1260
QY 1261 acaaacgatggagagatcagcaaaagacatgatggagcttgccaaatttataagaagt 1320
Db 1261 ACAACGATGGAGAGTAGCGCAAGAGCCATGATGGGACTTGGCCCAATTTATTAAGAAAT 1320
QY 1321 atgctttacagtcagcagctggaagaagatgctgcaaacagatagcatggatcgaagaca 1380
Db 1321 ATGCTTTACAGTCAGCAGCTGGAAAGATGCTGCAAAACAGATAGCATGGATCAAGACA 1380
QY 1381 aatgctacatatattatcaaatagattgtatgatcgactctgttgaacggatct 1440
Db 1381 ANTGTCTACATATATATATCAAAATAGTATTGATGATCGACTACTTGTGTGAACGGATCT 1440
QY 1441 ttgtcctaatacatgttctcctaacaatttagaaactcagaaactcagaaactcagaaact 1500
Db 1441 TTGCTCAATACATGTTCTCCACAAATTTAGNAACTACAGAACGGATGAAATGCTTATATT 1500
QY 1501 acttgatgcacactggatttaaatgctgtgaagcattgaaatggaatgtgc 1560
Db 1501 ACTTGTATGCCACACTCGATTTAAATGCTGTGAAGCATTTGAATGAAATGTGGAATGTC 1560
QY 1561 aaaaactgtccagacatcaagtaaaagattgcttcaacttgatgaagcaaaccaaacag 1620
Db 1561 AAATCTGCTCCGACATCAAGTAAGGATTTGCTTGACTTTGATTAAGCAACCCCAAAACAG 1620
QY 1621 atgccagtgtaaggccatattttcaaaagtgtatggttattacaagaaattttacctgac 1680
Db 1621 ATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAGTTTACCTGATC 1680
QY 1681 ctggttaaggtcaggtattcatgaagaaattcacacaggtgttagaagatgatgagaaaa 1740
Db 1681 CTGGTAAGGCTCAGGATTTTCATGAAGAAATTCACAGAGTGTGTAGAAGATGATGAGAAAA 1740
QY 1741 taagaagcaggttagaagtacttgttagtccaacatgctcctcgaagcaggtgtgaaggtt 1800
Db 1741 TAAGAAAGCAGTTAGAAGTACTTGTAGTCCAAACATGCTCCTGCAAGCAGGCTGAAGGTT 1800
QY 1801 gttggtgaaataactaagaaggttgggcaaccccccaacagactcaaaatcttctctgg 1860
Db 1801 GTGTGGGTGAAATTAACATAAGAGTTGGGCAACCCCAACAGGCTACAAATCTCTTCTGG 1860
QY 1861 aaatgatcaagttctcttggagagatagcacctgtgcacatagataccgaatctatca 1920
Db 1861 AAATGATCAAGTTTCTTTGGAGAGATAGCACCTGTGCACATAGATACCGAATCTATCA 1920
QY 1921 gtgctcttataaacaagtgaacaaatcaatagatggaacacagatgatgagatgaggg 1980
Db 1921 GTGCTCTTATTAACAAGTGAACAAATCAATAGATGGAACACACAGATGATGAAGATGAGG 1980

QY 1981 gtgttcoaatgatcaagcccatcagagcaggtcttgaactgttaaagtaactctcattta 2040
Db 1981 GTGTTCCAACATGATCAAGCCATCAGAGCGTCTTGAACACTCTTAAGTACTCTCAATTTA 2040
QY 2041 cacatcccatctcatttcattctgtctgaacatttgaatcattactgtgtgtctgaaaa 2100
Db 2041 CACATCCCATCTCATTTCAATCTGCTGAAACATTTGAATCATTTACTGGCCTGTCTGAAA 2100
QY 2101 tggatgatgaaaaagtaacagagaagctgacactacaaaattttcaaaaaacacaggaagcaaa 2160
Db 2101 TGGATGATGAAAAAGTAGCAAGAGCTGCATACAAATTTTCAAAAACACAGGAAGCAAAA 2160
QY 2161 ttgaagagattttccacacatcagatcagcctgtgctcctgttttacatcacaaatcta 2220
Db 2161 TTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCTGTTTACATCACAAATCTA 2220
QY 2221 aaaaagagaccccccgctcaagcacaatatccattcattgtatccatgcgatatcttcta 2280
Db 2221 AAAAGGACCCCCCGCTCAAGCCCAAAATATGCCATTTCAATTCATCCATGCCATATTTTCTA 2280
QY 2281 gtaaaagagaccccgcttgcacagatatgtgacacatttggctactattgtctcctctgcacctg 2340
Db 2281 GTAAAGAGGCCAGTTTGCACAGATATTGAGGCTCTGCAATAAGAGCCTAGATCCAAAGCA 2340
QY 2341 acctggaacatctcotaacaacacatttggctactattgtgtcatattgtctcctctgcacctg 2400
Db 2341 ACCTGGAACATCTCATAAACACCATTTGTTACTATTGGTCATATTGCTCTCTCTCTGACCTG 2400
QY 2401 atcaatttgcctcctctggaaaactcttggtagctactcttcatgttgaagatctctctca 2460
Db 2401 ATCAATTTGCTGCTCTCTTTGAAATCTTTGCTAGTACTTTTCAATTTGAAGAGATCTTCTCA 2460
QY 2461 tgaatgatcggtctccagggaaaaagacaactaaaccttgggttccagatgaagaagtat 2520
Db 2461 TGAATGATCGGCTTCCAGGGAAAAACACAACTAAACTTTGGGTTCCAGATGAAGAAAGTAT 2520
QY 2521 ctctcgagacaattggtcacaatttcaggctattaaaaatgattggttcgacttctgtaa 2580
Db 2521 CTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAAATGATGTTGCTGATGGCTACTTTGGAA 2580
QY 2581 tgaataataatcacagtaaatcaggaactctcacttaagattgctcaacaacaattatgc 2640
Db 2581 TGAATAATTAATCACAGTAATCAGGAACCTTACCTTAAAGATTGCTTAACAAACAATATTGC 2640
QY 2641 atagtgtgagacttgacagaaacaggggaaaaatttagtaaacacagatatgtcacgctga 2700
Db 2641 ATAGTGATGGAGACTTGACAGAAACAGGGGAAAAATTAGTAAACACAGATATGTCCAGCTGA 2700
QY 2701 gacttgcctggtgagtgctatttgaagctggcacaagaacccctgttaccatgaataca 2760
Db 2701 GACTTGTCTGGGAGTGTCTATTGTGAAGCTGGCACAAAGAACCCCTGTACCATTGAATCA 2760
QY 2761 tcacattagaacaatatcagctatgtcatttagctatcaacgatgaatgctatcaagtaa 2820
Db 2761 TCACATTAGAACAATATCAGCTATGTGCTATTAGCTATCAACGATGAATGCTATCAAGTAA 2820
QY 2821 gacaagttgtgccagaaaacttcaaaagccttcccgctttacggttccacttgcagt 2880
Db 2821 GACAAGTGTGTTGCCAGAAAACCTTCAAAAGGCTTTCCCGTTTACGGCTTCCACTTGAGT 2880
QY 2881 atatggcaatctgtgccccttggcacaagaactcctgtaaaaagagagaagactcatgcta 2940
Db 2881 ATATGCGCAATCTGTGCCCTTTGTGCAAAAGATCTCTGTAAGAGAGAGAGAGCTCATGCTA 2940
QY 2941 ggcaatgtttggtgaaaaataataaataagcgcgagatctctgaagcagcagcagctg 3000
Db 2941 GGCAATGTTGTTGTAAGAAATATAAATGTGAAGCGGGAGTATCTGAAGCAGGATGCAGCTG 3000
QY 3001 ttagtgaataattattgtctcttaccagagatgtgtgttcccatatacaatccacttt 3060
Db 3001 TTAGTGAATAATTTATTTGTCTCTTACACAGATGTTGTTTCCCATATACAAATTCACCTTT 3060
QY 3061 tggcacatgacccagatttatgtcaaaagtacaggatatattgaacaacttaagaagttaaag 3120

[illegible]

| | | | | | | |
|------|-------|-----|--------|----|----------|-------------|
| 9 | 497.4 | 9.4 | 54398 | 73 | AC068224 | Homo sapi |
| c 10 | 443 | 8.4 | 72157 | 64 | AC016449 | Homo sapi |
| c 11 | 420.8 | 8.0 | 54398 | 73 | AC068224 | Homo sapi |
| c 12 | 399 | 7.6 | 340969 | 81 | AL512630 | Mus muscu |
| c 13 | 232.2 | 4.4 | 2212 | 89 | AK021757 | Homo sapi |
| c 14 | 225.4 | 4.3 | 72157 | 64 | AC016449 | Homo sapi |
| c 15 | 206.8 | 3.9 | 113704 | 92 | HS267P19 | Homo sapi |
| c 16 | 206.8 | 3.9 | 163642 | 79 | AL353724 | Homo sapi |
| c 17 | 179.6 | 3.4 | 195380 | 80 | AL358892 | Homo sapi |
| c 18 | 163 | 3.1 | 163642 | 79 | AL353724 | Homo sapi |
| c 19 | 149 | 2.8 | 150 | 54 | G43328 | WIAF-1982-S |
| c 20 | 149 | 2.8 | 150 | 54 | G43329 | WIAF-1983-S |
| c 21 | 131 | 2.5 | 145450 | 73 | AC068352 | Homo sapi |
| c 22 | 131 | 2.5 | 160962 | 67 | AC022862 | Homo sapi |
| c 23 | 131 | 2.5 | 167671 | 77 | AC090081 | Homo sapi |
| c 24 | 124 | 2.4 | 69740 | 73 | AC087748 | Homo sapi |
| c 25 | 121.4 | 2.3 | 145450 | 73 | AC068352 | Homo sapi |
| c 26 | 121.4 | 2.3 | 162775 | 67 | AC023006 | Homo sapi |
| c 27 | 119.2 | 2.3 | 173613 | 4 | AC007475 | Homo sapi |
| c 28 | 119.2 | 2.3 | 192763 | 4 | AC007475 | Homo sapi |
| c 29 | 119.2 | 2.3 | 192763 | 4 | AC007475 | Homo sapi |
| c 30 | 119.2 | 2.3 | 192763 | 4 | AC007475 | Homo sapi |
| c 31 | 115.4 | 2.2 | 262731 | 5 | AE003823 | Drosophila |
| c 32 | 108 | 2.0 | 1489 | 89 | AC087748 | Homo sapi |
| c 33 | 99.8 | 1.9 | 72118 | 10 | I66494 | Sequence 14 |
| c 34 | 98.6 | 1.9 | 162775 | 67 | AC023006 | Homo sapi |
| c 35 | 87.6 | 1.7 | 167671 | 77 | AC090081 | Homo sapi |
| c 36 | 86.6 | 1.6 | 153665 | 78 | AL138818 | Homo sapi |
| c 37 | 83.8 | 1.6 | 160962 | 67 | AC022862 | Homo sapi |
| c 38 | 80.2 | 1.5 | 184535 | 84 | CNS057CJ | Homo sapi |
| c 39 | 72.6 | 1.4 | 741119 | 71 | AC036177 | Homo sapi |
| c 40 | 72 | 1.4 | 1084 | 53 | CNS06PLB | Homo sapi |
| c 41 | 71.2 | 1.4 | 978 | 53 | CNS06PJK | Homo sapi |
| c 42 | 70.8 | 1.3 | 174384 | 61 | AC009524 | Homo sapi |
| c 43 | 70.6 | 1.3 | 66288 | 78 | AC090825 | Homo sapi |
| c 44 | 70.6 | 1.3 | 156550 | 64 | AC015830 | Homo sapi |
| c 45 | 70 | 1.3 | 131271 | 64 | AC015927 | Homo sapi |

| | | | | | |
|------------|--|------------|------|-----|-------------|
| RESULT 1 | HSU95825 | 5253 bp | mRNA | PRI | 05-APR-1999 |
| LOCUS | Human androgen-induced prostate proliferative shutoff associated | | | | |
| DEFINITION | protein (AS3) mRNA, complete cds. | | | | |
| ACCESSION | U95825 | | | | |
| VERSION | U95825.2 | GI:4559409 | | | |
| KEYWORDS | human. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| AUTHORS | Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo. | | | | |
| TITLE | 1 (bases 1 to 5253) | | | | |
| JOURNAL | Geck, P., Szelei, J., Jimenez, J., Soto, A.M. and Sonnenschein, C. | | | | |
| REFERENCE | Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996) | | | | |
| AUTHORS | 2 (bases 1 to 5253) | | | | |
| TITLE | Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M. | | | | |
| JOURNAL | Early gene expression during androgen-induced inhibition of | | | | |
| MEDLINE | proliferation of prostate cancer cells: a new suppressor candidate | | | | |
| AUTHORS | on chromosome 13, in the BRCA2-Rb1 locus | | | | |
| REFERENCE | J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999) | | | | |
| TITLE | 99229875 | | | | |
| JOURNAL | 3 (bases 1 to 5253) | | | | |
| MEDLINE | Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M. | | | | |
| AUTHORS | Direct Submission | | | | |
| REFERENCE | Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University | | | | |
| TITLE | Medical School, 136 Harrison Avenue, Boston, MA 02111, USA | | | | |
| JOURNAL | 4 (bases 1 to 5253) | | | | |
| REFERENCE | Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M. | | | | |
| AUTHORS | Direct Submission | | | | |
| TITLE | | | | | |

| | |
|--|--|
| Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University | |
| Medical School, 136 Harrison Avenue, Boston, MA 02111, USA | |
| Sequence update by submitter | |
| On Apr 5, 1999 this sequence version replaced gi:4539617. | |
| Location/Qualifiers | |
| 1. 5253 | |
| /organism="Homo sapiens" | |
| /db_xref="taxon:9606" | |
| /chromosome="13" | |
| /map="13q12-q13" | |
| /cell_line="LNCap" | |
| /cell_type="cancer cells" | |
| /tissue_type="prostate" | |
| /notes="exon 1 maps to p1 Artificial Chromosome PAC26H23; | |
| exons 2-24 map to cosmid 267p19; exons 25-34 map to p1 | |
| Artificial Chromosome PAC 49J10" | |
| 1. 5253 | |
| /gene="AS3" | |
| 1. 46 | |
| /gene="AS3" | |
| /number=1 | |
| 47. 173 | |
| /gene="AS3" | |
| /number=2 | |
| 66. 4241 | |
| /gene="AS3" | |
| /codon_start=1 | |
| /product="androgen-induced prostate proliferative shutoff | |
| associated protein" | |
| /protein_id="AAD22134.2" | |
| /db_xref="GI:4559410" | |
| /translation="MAHSKTRTNDGKITPPGVKEISDKISKEEMVRRLKMWVTFMD | |
| MDQDEEEKELYNLALHLASDFGLKHPDKOVRLVACCLADIFRIYAPETSPDK | |
| LKDIEMFTIROLKLEDTKSPQNFYFYLLENIAWVSTNICFELEDSEIETQIYRT | |
| LFVYLNHGNQVYHMHVMDLMSIIICEDGTVSQELLDVNLVPAHNNKQAYDLA | |
| KALLAKTAIEPYITNFNFQVLMGKTSISDLSEHVSLLILEYNDHSHLLSLVLPQ | |
| LEFKLSNDNEERLQVVKLLAKMFGAKDSLEASONKQCQVGLGRFNDIHPVRLCY | |
| VFASHLMNHDPDLKDLTEYLKVRSHDPPEAIRDHVIIVSIVTAAKDKILLVNDHLSN | |
| VGERTLDRWRVRKEAMGLAQIYKKYALQSAAGDAKQIAWKDKILLHLYONSID | |
| DRLLVERIFAQVMVPHNLETTERMKCLYLYATLDLNAVKALNEMWKNQNLRHQVKD | |
| LLDLIKOPKTDASVKAIFSKVMVITRSLPDKGAKDFMKFTQVDEDEKIRKQLEVL | |
| VSPSCSKAQECVREITKNGNKPQNPFLMIKFLLEIAPIVHIDTESISALIKQ | |
| VKNSIDGADDEGEVPTDQAIRAGLELLKLVLSHSTHPTISFHSAPFESLLACLKMSK | |
| KVAEALQIFKNTGSKIIEEDPHITRSALLPVHLHKSPPQPAALYHICHTAIFSSK | |
| EAQAFIPEPLHKSLLDNLPHLITPLVTIGHIALPADPQAPLKSILVATFIVKDLL | |
| MNDRLPGKTTKLVPPDEEVSPETWVKIQAIKMWVRLGLGNHSHKSGTSLRLIT | |
| ILHSDGLTEOGKISKSPMSRLRSLAAGSAIVKLAQEPYHEITLQEQALCALINDE | |
| CYQVQVPAQLHKLGLSLRLPLFYMALCALCAPKVPKERRAHARQCLVKNVREY | |
| LQRAVSEKLLSLLEPYVVPYTHLLAHPDYVKVQDIEQLKDVKECLMFLVLELMA | |
| KNNNSHAFIRKMWENTKQTKDAQGDDAKNNEKLYTVCDVAMNIIKSKTYSLES | |
| KDPVLPARFTQDPKNTKNTKNTKNTKNTKNTKNTKNTKNTKNTKNTKNTKNTKNTK | |
| TKSSRMETVGNASSSSSPGRIRKGRDSDSDHSENYEDYTMSSPLPGKSSDRDSD | |
| DLVSELEKPRGKTKPTVTEEEKLGMDLLTKLVQEQRPKGSORSRKRGHTASSEDQ | |
| QWPEEKRLKLENEDEONSPKKGKGRPPKPLGGTPEEPTMTKSKGSKKSG | |
| PPAPEEEEEERQSQGTEQKSKQHRVRRRAQRAESESSESTESTQSTQSTQSTQSTQ | |
| SKTSPSPKPNV" | |
| 174. .377 | |
| /gene="AS3" | |
| /number=3 | |
| 378. 464 | |
| /gene="AS3" | |
| /number=4 | |
| 465. .562 | |
| /gene="AS3" | |
| /number=5 | |
| 563. 689 | |
| /gene="AS3" | |
| /number=6 | |
| 690. .770 | |
| /gene="AS3" | |
| /number=7 | |
| 771. 911 | |
| /gene="AS3" | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 25, 2001, 14:27:17 : Search time 12186.1 Seconds
(without alignments)
6690.424 Million cell updates/sec

Title: US-09-512-581-1
Perfect score: 5271
Sequence: 1 ccgagagcccgagtgag.....gtaaaaaaaaaaaaaa 5271

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl: *
- 1: gb_ba1:*
 - 2: gb_ba2:*
 - 3: gb_ba3:*
 - 4: gb_in1:*
 - 5: gb_in2:*
 - 6: gb_in3:*
 - 7: gb_om:*
 - 8: gb_ov:*
 - 9: gb_pat1:*
 - 10: gb_pat2:*
 - 11: gb_ph:*
 - 12: gb_pl1:*
 - 13: gb_pl2:*
 - 14: gb_pl3:*
 - 15: gb_pl4:*
 - 16: em_ba1:*
 - 17: em_ba2:*
 - 18: em_fun:*
 - 19: em_htgo_hum:*
 - 20: em_htgo_inv:*
 - 21: em_htgo_rod:*
 - 22: em_htg_hum1:*
 - 23: em_htg_hum2:*
 - 24: em_htg_hum3:*
 - 25: em_htg_hum4:*
 - 26: em_htg_hum5:*
 - 27: em_htg_hum6:*
 - 28: em_htg_hum7:*
 - 29: em_htg_hum8:*
 - 30: em_htg_inv1:*
 - 31: em_htg_inv2:*
 - 32: em_htg_other:*
 - 33: em_htg_rod:*
 - 34: em_hum1:*
 - 35: em_hum2:*
 - 36: em_hum3:*
 - 37: em_hum4:*
 - 38: em_hum5:*
 - 39: em_hum6:*
 - 40: em_hum7:*
 - 41: em_in:*
 - 42: em_om:*
 - 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vil2:*

60: gb_vil3:*

61: gb_vil4:*

62: gb_vil5:*

63: gb_vil6:*

64: gb_vil7:*

65: gb_vil8:*

66: gb_vil9:*

67: gb_vil10:*

68: gb_vil11:*

69: gb_vil12:*

70: gb_vil13:*

71: gb_vil14:*

72: gb_vil15:*

73: gb_vil16:*

74: gb_vil17:*

75: gb_vil18:*

76: gb_vil19:*

77: gb_vil20:*

78: gb_vil21:*

79: gb_vil22:*

80: gb_vil23:*

81: gb_vil24:*

82: gb_vil25:*

83: gb_vil26:*

84: gb_vil27:*

85: gb_vil28:*

86: gb_vil29:*

87: gb_vil30:*

88: gb_vil31:*

89: gb_vil32:*

90: gb_vil33:*

91: gb_vil34:*

92: gb_vil35:*

93: gb_vil36:*

94: gb_vil37:*

95: gb_vil38:*

96: gb_vil39:*

97: gb_vil40:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|--------|---------|--------------|-------------|---------------------|
| 1 | 5221 | 99.1 | 5253 | 97 HS095825 | U95825 Human andro |
| 2 | 5084.4 | 96.5 | 7444 | 92 HS26H231 | AL137201 Novel hum |
| 3 | 4906.2 | 93.1 | 5110 | 85 AB023196 | AB023196 Homo sapi |
| 4 | 1754.4 | 33.3 | 1852 | 97 HS050533 | U050533 Human BRCA2 |
| 5 | 1041.4 | 19.8 | 5177 | 85 AB014548 | AB014548 Homo sapi |
| 6 | 901.6 | 17.1 | 137246 | 92 HS49J10 | Z84572 Human DNA s |
| 7 | 877.6 | 16.6 | 168487 | 89 AL138820 | AL138820 Human DNA |
| 8 | 566.4 | 10.7 | 2079 | 89 AK026889 | AK026889 Homo sapi |

J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK

TITLE JOURNAL COMMENT

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1946 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 390.

FEATURES

Location/Qualifiers
1..391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:856164"
/clone_lib="Stratagene HeLa cell s3 937216"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTT 3' "

BASE COUNT 112 a 81 c 64 g 133 t 1 others
ORIGIN

alignment_scores:

Quality: 667.00 Length: 129
Ratio: 5.171 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.225

alignment_block:

US-09-512-581-2 x AA630616/rev ..

Align seg 1/1 to reverse of: AA630616 from: 1 to: 391

```

930 LeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCy 946
|||||
389 CTTTCCCGTTTACGGCTTCCACTTGAGTATATGCGAATCTGCGCCTTG 340
|||||
946 sAlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuV 963
|||||
339 TGCAAAAGATCCTGTAAGGAGAGAGAGAGCTCATGCTAGGCAATGTTGG 290
|||||
963 alLysAsnIleAsnValArgGluTyrLeuLysGlnHisAlaAlaVal 979
|||||
289 TGAATAATATAAATGTAAGCGGGAGTATCTGAAGCAGCATGCAGCTGTT 240
|||||
980 SerGluLysLeuLeuSerLeuLeuProGluTyrValProTyrThrII 996
|||||
239 AGTGAAATAATTATTGCTCTCTACACAGATATGTTCTTCATATACAA 190
|||||
996 eHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluG 1013
|||||
189 TCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGATATTGAAC 140
|||||
1013 lnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMet 1029
|||||
139 AACTTAAAGATGTTAAAGAATGCTTTGGTTTGTCTTGGAAATATTAATG 90
|||||
1030 AlaLysAsnGluAsnSerHisAlaPheIleArgLysMetValGluAs 1046
|||||
89 GCTTAAATGAAATACAGTCACGCTTTTATCAGAAAGATGCTAGAAA 40
|||||
1046 nIleLysGlnThrLysAspAlaGlnGlyProAspAsp 1058

```

|||||
39 TATTAACAACAACAAAAGATGCCCAAGGACCAGATGAT 3

THIS PAGE BLANK (USPTO)